Politechnika Wrocławska
Wydział Inżynierii Środowiska
Katedra Biologii Środowiskowej i Ochrony Atmosfery

PRACA DOKTORSKA

Prognozowanie stężeń zanieczyszczeń powietrza z wykorzystaniem metod uczenia głębokiego

Jarosław Bernacki

Promotor: dr hab. inż. Izabela Sówka, prof. uczelni

Wrocław, 2026

Streszczenie pracy

Zanieczyszczenie powietrza stanowi jedno z kluczowych zagrożeń środowiskowych, szczególnie w krajach o wysokim poziomie uprzemysłowienia i gęstej zabudowie miejskiej. Skuteczne prognozowanie stężeń pyłów zawieszonych, zwłaszcza frakcji $PM_{2,5}$ i PM_{10}, ma istotne znaczenie dla analiz stanu atmosfery oraz wspomagania działań podejmowanych przez instytucje odpowiedzialne za zarządzanie jakością powietrza.

W niniejszej pracy zaproponowano i zaimplementowano nowoczesne podejścia oparte na metodach głębokiego uczenia. Wykorzystano różne topologie sieci konwolucyjnych oraz autoenkoderów do prognozowania krótkoterminowych stężeń zanieczyszczeń powietrza. W badaniach przyjęto podejście czasowe, w którym prognozy krótkoterminowych stężeń zanieczyszczeń uzyskiwano na podstawie wartości historycznych (ang. *data-driven*). Modele zostały wytrenowane i przetestowane na rzeczywistych danych pomiarowych pochodzących z automatycznych stacji monitoringu jakości powietrza zlokalizowanych w różnych regionach Polski. Wykorzystano informacje meteorologiczne oraz sezonowe, co pozwoliło na lepsze odwzorowanie zależności między zmiennymi środowiskowymi. W badaniach położono nacisk na ocenę efektywności proponowanych modeli w zróżnicowanych warunkach środowiskowych Polski, co w literaturze światowej jest rzadko analizowane w tak szerokiej skali regionalnej.

Uzyskane wyniki wykazały, że modele głębokiego uczenia, zwłaszcza z architekturą konwolucyjną, osiągają lepsze parametry dokładności niż klasyczne podejścia statystyczne czy sieci neuronowe o płaskiej strukturze. Dokładność metod oceniono przy użyciu standardowych miar błędu (MAE, MSE, RMSE, R^2) oraz testów statystycznych. Analiza rezultatów wykazała, że proponowane metody uzyskały niższe wartości miar błędów predykcji oraz wyższe wartości dopasowania R^2. Analiza hipotez statystycznych również potwierdziła przewagę proponowanych metod w odniesieniu do metod z literatury.

Przeprowadzone badania potwierdzają zasadność wykorzystania metod głębokiego uczenia w zadaniu krótkoterminowego prognozowania jakości powietrza w warunkach silnego zróżnicowania przestrzennego. Zaproponowane podejście umożliwia ocenę skuteczności modeli prognostycznych w skali obejmującej wiele lokalizacji o odmiennych uwarunkowaniach środowiskowych, co stanowi istotne rozszerzenie dotychczasowych badań prowadzonych najczęściej w ograniczonych warunkach lokalnych. Uzyskane wyniki można uznać za punkt wyjścia do dalszego rozwoju zaproponowanych metod w kierunku bardziej złożonych analiz czasowych oraz przestrzennych. Jednym z możliwych kierunków dalszego rozwoju jest rozszerzenie aspektu prognozowania do podejścia czasowo-przestrzennego, w którym zakłada się uwzględnienie informacji historycznych z sąsiednich stacji pomiarowych.

Abstract

Air pollution is an environmental threat, particularly in countries with a high level of industrialization and dense urban development. Effective forecasting of particulate matter concentrations, especially fractions $PM_{2,5}$ and PM_{10}, is crucial for analyzing the state of the atmosphere and supporting actions taken by institutions responsible for air quality management.

This dissertation proposes and implements modern approaches based on deep learning methods. Various topologies of convolutional neural networks and autoencoders were used to forecast short-term air pollutant concentrations. The research adopted a time-based approach, in which forecasts of short-term pollutant concentrations were obtained based on historical values (data-driven approach). The models were trained and tested on real-world measurement data from automatic air quality monitoring stations located in various regions of Poland. Meteorological and seasonal information were used, allowing for a better representation of the relationships between environmental variables. The study focused on assessing the effectiveness of the proposed models under Poland's diverse environmental conditions, which have rarely been analyzed at such a broad regional scale in the existing literature.

The obtained results showed that deep learning models, especially those with a convolutional architecture, achieve better accuracy parameters than classical statistical approaches or flat neural networks. The accuracy of the methods was assessed using standard error measures (MAE, MSE, RMSE, R^2) and statistical tests. Analysis of the results showed that the proposed methods achieved lower prediction error measures and higher goodness-of-fit values R^2. Analysis of the statistical hypotheses also confirmed the advantages of the proposed methods compared to those from the literature.

The conducted research confirms the validity of using deep learning methods for short-term air quality forecasting in conditions of strong spatial variation. The proposed approach enables the assessment of the effectiveness of forecasting models on a scale encompassing multiple locations with different environmental conditions, which represents a significant extension of previous research, which was usually conducted in limited local conditions. The obtained results can be considered a starting point for further development of the proposed methods towards more complex temporal and spatial analyses. One possible direction for further development is to expand the forecasting aspect to a temporal-spatial approach, which assumes the inclusion of historical information from neighboring measurement stations.

Spis treści

Wykaz skrótów i oznaczeń

AI	sztuczna inteligencja
ANFIS	adaptacyjny neuro-rozmyty system wnioskowania
ANN	sztuczna sieć neuronowa
API	indeks zanieczyszczenia powietrza
AQHI	wskaźnik zdrowia jakości powietrza
AQI	indeks jakości powietrza
AR	model autoregresyjny
ARIMA	zintegrowany model autoregresyjno-średniej ruchomej
ARMA	model autoregresyjno-średniej ruchomej
As	arsen
BaP	benzo(a)piren
Bi-LSTM	dwukierunkowa długa pamięć krótkotrwała
C_6H_6	benzen
CAE	autoenkoder konwolucyjno-rekurencyjny
CAI	kompleksowy indeks jakości powietrza
CAQI	powszechny wskaźnik jakości powietrza
Cd	kadm
CEEMD	kompletna zespołowa empiryczna dekompozycja trybów
CNN	konwolucyjna sieć neuronowa
CO	tlenek węgla(II)
CO_2	tlenek węgla(IV)
CPM	pył gruby
DL	uczenie głębokie
DT	drzewo decyzyjne
EAQ	Europejski System Monitoringu Jakości Powietrza
EEMD	zestawowa dekompozycja empiryczna
EMD	empiryczna dekompozycja sygnału
ESN	sieć neuronowa typu echo
FCN	sieć w pełni konwolucyjna
GAM	modele addytywne
GB	wzmocnienie gradientowe
GCN	grafowa sieć konwolucyjna
GIOŚ	Główny Inspektorat Ochrony Środowiska

GRU	bramkowa sieć rekurencyjna
HA	autoenkoder hierarchiczny
IAQI	indywidualny wskaźnik jakości powietrza
KAN	sieć Kolmogorow-Arnold
LMM	liniowy model mieszany
LR	regresja liniowa
LSTM	długa pamięć krótkotrwała
MAE	średni błąd bezwzględny
MAPE	średni procentowy błąd bezwzględny
MLP	perceptron wielowarstwowy
MLR	regresja wieloraka
MSE	błąd średniokwadratowy
MVO	optymalizacja średnio-wariancyjna
NARX	nieliniowy autoregresyjny model ze zmiennymi egzogenicznymi
Ni	nikiel
NO	tlenek azotu(II)
NO_2	tlenek azotu(IV)
NO_x	tlenki azotu
O_3	ozon
Pb	ołów
PM	pył zawieszony
PM_{10}	pył zawieszony o średnicy cząstek nie większej niż 10 μm
$PM_{2,5}$	pył zawieszony o średnicy cząstek nie większej niż 2,5 μm
PM_1	pył zawieszony o średnicy cząstek nie większej niż 1 μm
PMŚ	Państwowy Monitoring Środowiska
POChP	przewlekła obturacyjna choroba płuc
PSI	wskaźnik norm zanieczyszczeń
PSO	optymalizacja rojem cząstek
QR	regresja kwantylowa
R	współczynnik korelacji
R^2	współczynnik dopasowania
RBF	radialna funkcja bazowa
RF	las losowy
RMSE	pierwiastek błędu średniokwadratowego
RNN	rekurencyjna sieć neuronowa

SARIMA	sezonowy zintegrowany model autoregresyjno-średniej ruchomej
SO	tlenek siarki(II)
SO_2	tlenek siarki(IV)
SSA	algorytm poszukiwań wróbli
SVM	maszyna wektorów nośnych
SVR	regresja wektorów wspierających
TCN	konwolucyjna sieć temporalna
TFT	sieć typu *transformer* z fuzją czasową
TSP	całkowite pyły zawieszone
VAE	autoenkoder wariancyjny
VARMA	wektorowy model autoregresyjno-średniej ruchomej
WANN	falowe sieci neuronowe
WRF	model numeryczny prognozowania pogody
WT-ARIMA	model ARIMA z transformacją falkową
xLSTM	rozszerzona sieć z długą pamięcią krótkotrwałą

Rozdział 1

Wstęp

Zanieczyszczenie powietrza stanowi jeden z najpoważniejszych globalnych problemów środowiskowych, mających istotny wpływ na zdrowie ludzi oraz funkcjonowanie ekosystemów. Jakość powietrza jest kluczowym czynnikiem determinującym stan zdrowia populacji, ponieważ zanieczyszczone powietrze zawiera liczne substancje szkodliwe, takie jak pył zawieszony (ang. *particulate matter* – PM), tlenki azotu czy benzo(a)piren. Związki te mogą prowadzić do rozwoju poważnych chorób układu oddechowego i krążenia [9, 210]. Długotrwała ekspozycja na zanieczyszczenia powietrza zwiększa ryzyko wystąpienia astmy, przewlekłej obturacyjnej choroby płuc (POChP) oraz nowotworów [28, 137]. Szczególnie narażone na negatywne skutki są dzieci, osoby starsze oraz osoby cierpiące na choroby przewlekłe [76].

Stężenia zanieczyszczeń powietrza, w szczególności pyłów zawieszonych, przyczyniają się do znacznej liczby przedwczesnych zgonów na całym świecie. W wielu regionach Europy jakość powietrza pozostaje niezadowalająca, zwłaszcza na obszarach silnie zurbanizowanych [162]. Za główną przyczynę zanieczyszczenia powietrza można uznać procesy spalania paliw kopalnych, prowadzące do emisji takich substancji jak benzen (C_6H_6), tlenek węgla(II) (CO), tlenek węgla(IV) (CO_2), tlenki azotu (NO, NO_2, NO_x), pyły zawieszone (PM_1, $PM_{2,5}$, PM_{10}) [135] oraz dwutlenek siarki (SO_2). Z tego względu jakość powietrza jest przedmiotem systematycznego monitoringu, a jej ocena opiera się m.in. na analizie danych pochodzących z sieci stacji pomiarowych.

Według Światowej Organizacji Zdrowia (WHO), spośród 50 miast Unii Europejskiej o najgorszej jakości powietrza aż 33 znajdują się na terytorium Polski [22, 98]. Pomimo obserwowanej w ostatnich latach redukcji emisji pyłów zawieszonych, nadal notowane są wysokie stężenia frakcji pyłu zawieszonego $PM_{2,5}$ oraz PM_{10}, szczególnie w sezonie grzewczym. Jednocześnie jakość powietrza wykazuje istotne zróżnicowanie regionalne [187]. Najbardziej zanieczyszczone obszary obejmują południową część kraju, w tym przede wszystkim region Górnego Śląska, natomiast relatywnie dobrą jakość powietrza obserwuje się w północnej Polsce, w rejonach nadmorskich oraz na wybranych obszarach Ziemi Lubuskiej [119, 208].

Systematyczna ocena stanu jakości powietrza w Polsce realizowana jest w ramach Państwowego Monitoringu Środowiska (PMŚ), który stanowi podstawowe narzędzie gromadzenia, przetwarzania oraz udostępniania danych o stanie środowiska. W ramach PMŚ prowadzone są ciągłe i okresowe pomiary stężeń substancji zanieczyszczających powietrze, w tym pyłów zawieszonych frakcji $PM_{2,5}$ i PM_{10}, co umożliwia analizę długoterminowych trendów, ocenę narażenia ludności oraz identyfikację obszarów problemowych [159].

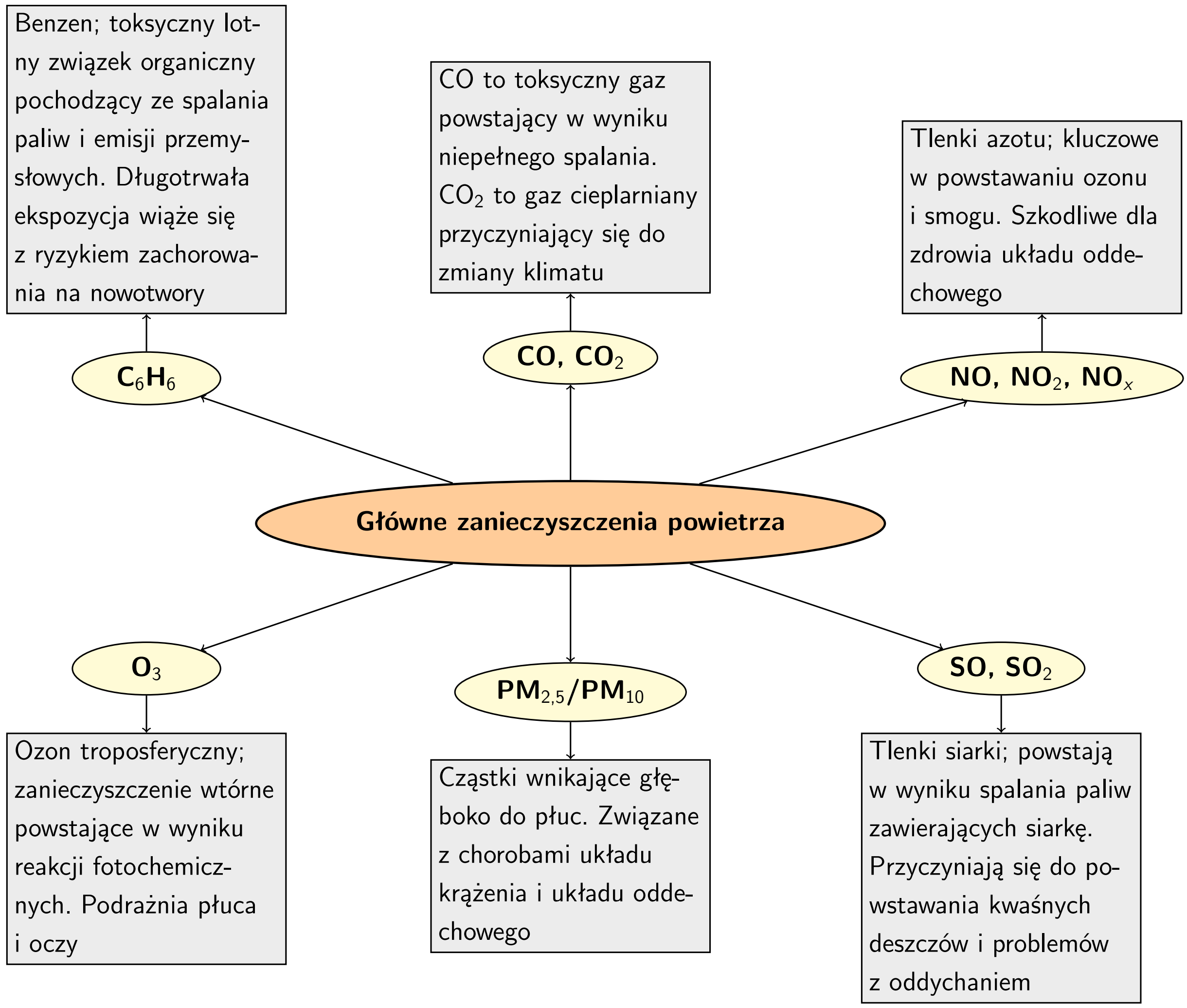
Instytucją odpowiedzialną za realizację zadań PMŚ w zakresie monitoringu jakości powietrza jest Główny Inspektorat Ochrony Środowiska (GIOŚ) [74]. GIOŚ koordynuje funkcjonowanie krajowej sieci stacji pomiarowych, odpowiada za weryfikację i publikację wyników pomiarów oraz zapewnia ich zgodność z obowiązującymi kryteriami oceny jakości powietrza, określonymi w odpowiednich aktach prawnych. Dane pochodzące z monitoringu prowadzonego przez GIOŚ stanowią podstawę do wczesnego ostrzegania przed epizodami smogowymi oraz do podejmowania działań mających na celu ograniczenie emisji zanieczyszczeń i poprawę jakości powietrza, w szczególności na obszarach zurbanizowanych [162].

Do klasyfikacji jakości powietrza oraz informowania społeczeństwa o poziomie zagrożenia zdrowotnego stosuje się indeks jakości powietrza (ang. *Air Quality Index* – AQI) [32] oraz indeks zanieczyszczenia powietrza (ang. *Air Pollution Index* – API) [190]. AQI, powszechnie stosowany na całym świecie, integruje informacje o kilku kluczowych zanieczyszczeniach i prezentuje wyniki w czytelnej skali liczbowej i kolorystycznej. API jest natomiast starszym wskaźnikiem, wykorzystywanym głównie w wybranych krajach azjatyckich. Oprócz nich stosowane są również inne indeksy, takie jak AQHI (ang. *Air Quality Health Index*) [191], CAQI (ang. *Common Air Quality Index*) [200], CAI (ang. *Comprehensive Air-quality Index*) [114], PSI (ang. *Pollution Standards Index*) [32] oraz IAQI (ang. *Indoor Air Quality Index*) [31], które różnią się zakresem monitorowanych zanieczyszczeń oraz sposobem interpretacji wyników. Zestawienie wybranych wskaźników jakości powietrza przedstawiono w tabeli 1.1.

Monitorowanie jakości powietrza obejmuje również prognozowanie stężeń zanieczyszczeń, które odgrywa istotną rolę dla skutków zdrowotnych i środowiskowych. Przykładowo, prognozowane są stężenia pyłów zawieszonych frakcji PM_{10} i $PM_{2,5}$, które ze względu na niewielkie rozmiary cząstek mogą przenikać do układu oddechowego i krwionośnego, powodując poważne konsekwencje zdrowotne [91]. W zależności od lokalnych warunków meteorologicznych oraz struktury źródeł emisji, prognozy mogą obejmować także inne substancje, takie jak benzo(a)piren czy lotne związki organiczne [175]. Główne zanieczyszczenia powietrza oraz ich krótką charakterystykę przedstawiono na rys. 1.1.

Tabela 1.1. Wybrane wskaźniki jakości powietrza stosowane w różnych krajach i regionach.

Indeks	**Kraj / Region**	**Monitorowane zanieczyszczenia**	**Skala**	**Uwagi**
AQI	USA, Chiny, Indie, inne kraje	$PM_{2,5}$, PM_{10}, O_3, CO, NO_2, SO_2	0–500	szeroko stosowane na całym świecie; kategorie zdrowotne od „dobrego" do „niebezpiecznego"
API	Malezja, Hongkong (przed 2013 r.)	PM_{10}, SO_2, NO_2, O_3, CO	0–500	zastąpione przez AQHI w Hongkongu w 2013 r.
AQHI	Kanada, Hongkong (od 2013 r.)	$PM_{2,5}$, O_3, NO_2	1–10+	skoncentrowane na ryzyku zdrowotnym
CAQI	Unia Europejska	$PM_{2,5}$, PM_{10}, NO_2, O_3	0–100+	wykorzystywane w mapach jakości powietrza
CAI	Korea Południowa	$PM_{2,5}$, PM_{10}, O_3, CO, NO_2, SO_2	0–500	obejmuje zarówno bieżące, jak i prognozowane warunki
PSI	Singapur	$PM_{2,5}$, PM_{10}, O_3, CO, NO_2, SO_2	0–500	na podstawie wskaźnika jakości powietrza w USA
IAQI	Chiny	każde zanieczyszczenie osobno (PM, gazy)	0–500	używane jako składniki ogólnego wskaźnika jakości powietrza



Rysunek 1.1. Główne zanieczyszczenia powietrza wraz z krótką charakterystyką [188].

Informacje o prognozowanej jakości powietrza odgrywają istotną rolę w planowaniu polityki środowiskowej, m.in. w podejmowaniu decyzji dotyczących ograniczeń ruchu samochodowego, wprowadzania alertów smogowych czy działań prewencyjnych. Rozwój nowoczesnych metod opartych na sztucznej inteligencji oraz modelach numerycznych umożliwia coraz dokładniejszą analizę i prognozowanie zmian jakości powietrza. Im wyższa dokładność prognoz, tym skuteczniejsze mogą być działania zapobiegawcze, co bezpośrednio przekłada się na poprawę zdrowia publicznego. Z tego względu rozwój i doskonalenie metod prognozowania jakości powietrza powinno stanowić jeden z priorytetów współczesnej polityki środowiskowej [204].

Niniejsza praca koncentruje się na prognozowaniu stężeń wybranych zanieczyszczeń powietrza, istotnych z punktu widzenia zdrowia publicznego i ochrony środowiska. W przeprowadzonych analizach wykorzystano wyniki pomiarów stężeń pyłów zawieszonych frakcji $PM_{2,5}$ i PM_{10}, a w wybranych wariantach badań uwzględniono również dane w zakresie stężeń gazowych zanieczyszczeń powietrza, takich jak CO, NO, NO_2, NO_x, SO_2 oraz O_3. Istniejące w literaturze metody prognozowania stężeń zanieczyszczeń powietrza są zwykle testowane na ograniczonej liczbie stacji pomiarowych, często zlokalizowanych w pojedynczych miastach lub regionach. Wąski zakres danych może ograniczać skuteczność proponowanych metod w zróżnicowanych warunkach środowiskowych, takich jak obszary o wysokim stopniu urbanizacji, regiony przemysłowe czy miejscowości uzdrowiskowe. W związku z tym, w niniejszej pracy podjęto próbę opracowania i oceny metod prognozowania krótkoterminowych stężeń pyłów zawieszonych i gazów zanieczyszczających w Polsce z wykorzystaniem danych (ang. *data-driven*) pochodzących z szerokiej sieci automatycznych stacji monitoringu jakości powietrza. Analiza obejmuje różnorodne lokalizacje geograficzne i warunki środowiskowe (wymienione w podrozdz. 1.2, co pozwala na kompleksową ocenę efektywności modeli w skali krajowej i stanowi punkt wyjścia do dalszych badań nad ich adaptacją do zmiennych warunków lokalnych. W proponowanych metodach zastosowano podejścia oparte na głębokim uczeniu, w tym sieci konwolucyjne i autoenkodery, co pozwala na lepsze odwzorowanie złożonych zależności w danych i stanowi punkt wyjścia do dalszych badań nad adaptacją modeli do zmiennych warunków lokalnych.

1.1 Aktualny stan badań

Podejścia do prognozowania zanieczyszczenia powietrza różnią się znacząco w poszczególnych regionach świata, odzwierciedlając różnorodność lokalnych warunków środowiskowych, dostępności danych i dominujących źródeł emisji. Na obszarach o silnym uprzemysłowieniu lub urbanizacji, takich jak części Azji czy Europy Wschodniej, modele prognostyczne często wykorzystują złożone architektury głębokiego uczenia (ang. *deep learning*), aby uchwycić nieliniowe zależności między wieloma zmiennymi. Z kolei regiony o ograniczonej infrastrukturze monitorującej mogą polegać na prostszych metodach statystycznych lub integrować dane satelitarne, aby kompensować luki w danych na poziomie gruntu. Wpływy meteorologiczne, cechy topograficzne i wzorce sezonowe również wpływają na stosowane strategie modelowania. W rezultacie systemy prognostyczne muszą być dostosowane do specyfiki regionalnej, aby zapewnić dokładność i praktyczną użyteczność w procesie decyzyjnym i ochronie zdrowia publicznego. Niniejsza sekcja zawiera przegląd metod prognozowania stężeń powietrza.

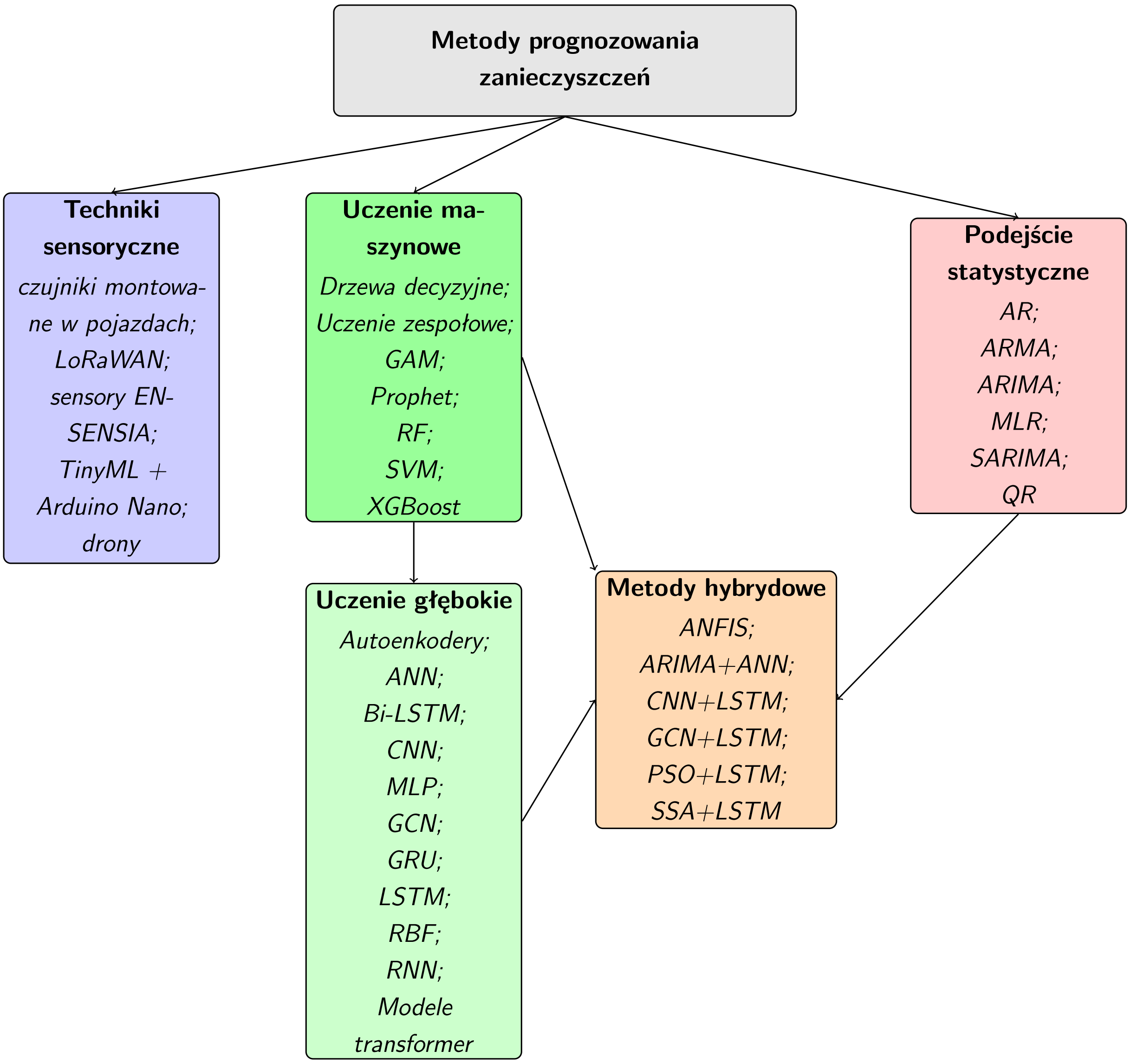
Prognozowanie stężeń zanieczyszczeń powietrza realizowane jest z wykorzystaniem szerokiego spektrum metod, które można podzielić na kilka głównych kategorii, jak przedstawiono na rys. 1.2 [142]. Do pierwszej grupy należą zaawansowane techniki pozyskiwania danych, obejmujące m.in. czujniki mobilne montowane na pojazdach, sieci sensorów opartych na technologii LoRaWAN [193], systemy niskokosztowych czujników środowiskowych [94, 131], rozwiązania klasy TinyML oraz platformy pomiarowe wykorzystujące bezzałogowe statki powietrzne (drony) [161]. Metody te umożliwiają zwiększenie rozdzielczości przestrzennej i czasowej danych wejściowych, stanowiąc istotne uzupełnienie klasycznych stacji monitoringu.

Kolejną grupę stanowią metody statystyczne, takie jak modele autoregresyjne (AR, ARMA, ARIMA, SARIMA), regresja wieloraka (MLR) oraz regresja kwantylowa (QR), które od dawna są stosowane w analizie szeregów czasowych oraz krótkoterminowym prognozowaniu stężeń zanieczyszczeń powietrza. Metody te charakteryzują się relatywnie prostą interpretacją oraz niewielkimi wymaganiami obliczeniowymi, jednak ich skuteczność bywa ograniczona w przypadku danych silnie nieliniowych oraz niestacjonarnych [41, 109, 138, 163, 198].

W ostatnich latach popularność zyskały metody uczenia maszynowego, w tym drzewa decyzyjne, metody zespołowe, lasy losowe (RF), maszyny wektorów nośnych (SVM), modele addytywne (GAM) oraz algorytmy wzmocnienia gradientowego, takie jak XGBoost [172, 128, 221, 230]. W ramach tej grupy szczególne znaczenie zyskują metody uczenia głębokiego (ang. *deep learning*), obejmujące m.in. sztuczne sieci neuronowe (ANN), sieci rekurencyjne (RNN), w tym LSTM, GRU i Bi-LSTM, konwolucyjne sieci neuronowe (CNN), grafowe sieci neuronowe (GCN) oraz modele oparte na mechanizmach uwagi (ang. *attention mechanism*) [203, 56, 194], w tym sieci typu transformer (ang. *transformers*) [46, 124, 220]. Metody te umożliwiają skuteczne modelowanie złożonych zależności czasowych i przestrzennych występujących w danych dotyczących jakości powietrza.

Kolejną grupę stanowią metody hybrydowe, które łączą podejścia statystyczne, algorytmy uczenia maszynowego oraz techniki optymalizacji. Do tej kategorii należą m.in. adaptacyjny neuro-rozmyty system wnioskowania (ang. *adaptive neuro-fuzzy inference system* – ANFIS) [164, 218], modele ARIMA wspomagane sieciami neuronowymi [51], architektury

CNN–LSTM oraz GCN–LSTM [215], a także rozwiązania wykorzystujące algorytmy optymalizacyjne, takie jak optymalizacja rojem cząstek (PSO) czy algorytm poszukiwań wróbli (SSA), stosowane do strojenia parametrów modeli neuronowych [86]. Podejścia hybrydowe umożliwiają wykorzystanie komplementarnych zalet poszczególnych metod i często prowadzą do poprawy dokładności prognoz w porównaniu z modelami jednorodnymi [188].



Rysunek 1.2. Kategoryzacja metod prognozowania z wykorzystaniem wybranych metod [188].

1.1.1 Metody prognozowania stężeń zanieczyszczeń powietrza stosowane na świecie ze szczególnym uwzględnieniem uczenia maszynowego

Zastosowanie sensorów Prowadzone na świecie badania pozwalają wyodrębnić główne kierunki wykorzystania sensorów w prognozowaniu jakości powietrza. Przykładowo, w pracy Zafra-Pérez i in. [223] zaproponowano niskokosztową bezprzewodową sieć LoRaWAN do pomiaru PM, przetestowaną w środowisku przemysłowym. Urządzenia pomiarowe oparto na czujnikach SDS011 oraz komponentach Arduino i Raspberry Pi. Przeprowadzono lokalną kalibrację, uzyskując błąd MAE poniżej 0,3 $\mu g/m^3$ oraz współczynnik determinacji $R^2 = 0{,}96$. Sieć umożliwiła zbieranie danych w czasie rzeczywistym i ich transmisję do platformy chmurowej. Z kolei w pracy Apostolopoulos i in. [7], celem badań była kalibracja niskokosztowych czujników ENSENSIA, mierzących stężenia NO_2, NO, O_3, CO oraz $PM_{2,5}$, w środowisku miejskim z wykorzystaniem metod uczenia maszynowego i głębokiego. Zebrane dane porównano z pomiarami ze stacji referencyjnych. Po kalibracji uzyskano wysoką zgodność wyników ($R^2 > 0{,}86$) oraz istotne zmniejszenie błędu pomiarowego. Podkreślono jednocześnie konieczność dynamicznej korekcji czujników klasy. W badaniu z pracy [90] zastosowano drony wyposażone w laserowe czujniki do detekcji gazów (C_6H_6, HCHO, SO_2) oraz pyłów zawieszonych (PM_1, $PM_{2,5}$, PM_{10}), a także moduły GPS, do monitorowania emisji liniowych, m.in. wzdłuż dróg i linii kolejowych. System umożliwiał analizę pionowego profilu zanieczyszczeń oraz ich rozkładu wzdłuż szlaków transportowych. Drony działały autonomicznie w zmiennych warunkach atmosferycznych, a wyniki porównano z danymi ze stacji referencyjnych.

W pracy [92] zaprezentowano system nazwany Enviro-IoT, który jest niskokosztowym, działającym w czasie rzeczywistym rozwiązaniem do monitorowania jakości powietrza, opartym na miniaturowych czujnikach oraz technologii Internetu Rzeczy. Autorzy opisali czteroletni proces projektowania i wdrażania systemu, a następnie przedstawili wyniki dziewięciomiesięcznych badań w środowisku miejskim. System porównano z przemysłową aparaturą referencyjną, uzyskując wysoką dokładność pomiarów frakcji pyłów zawieszonych $PM_{2,5}$, PM_{10} oraz NO_2 (odpowiednio 98%, 97% i 97%). Analiza ponad 57 000 próbek potwierdziła skuteczność rozwiązania, dowodząc, że niskokosztowe czujniki wspierane przez IoT mogą stanowić realną alternatywę dla tradycyjnych systemów pomiarowych. Badania opisane w [146] dotyczą opracowania niskokosztowego, samodzielnie skonstruowanego systemu oceny jakości powietrza, przeznaczonego do zastosowań wewnętrznych i zewnętrznych. System oparto na mikrokontrolerach ESP32 komunikujących się za pomocą protokołu ESP-NOW oraz czujnikach Alphasense do pomiarów aerozoli i gazów, takich jak NO_2, SO_2, O_3, CO, CO_2 i H_2S. Urządzenie wykorzystano do oceny powietrza zanieczyszczonego dymem tytoniowym po przejściu przez reaktor plazmowy DBD.

W pracy [39] skupiono się na opracowaniu uniwersalnej metody kalibracji niskokosztowych czujników jakości powietrza, umożliwiającej ich masowe wdrażanie w miejskich sieciach monitoringu. Autorzy zaproponowali globalną metodę kalibracji, niewymagającą próbek transferowych, opartą na danych z ograniczonej liczby czujników oraz algorytmach uczenia maszynowego. Testy przeprowadzone w różnych porach roku wykazały skuteczność porównywalną z indywidualną kalibracją każdego urządzenia. Zastosowanie tej metody może znacząco obniżyć koszty i umożliwić szerokie wykorzystanie precyzyjnych czujników IoT. Badanie [108], przeprowadzone w Delhi, dotyczyło analizy trzech popularnych czujników PM (SPS30, PMS7003 oraz HPMA115C0-004) w warunkach wysokiej wilgotności i zapylenia dla pomiarów frakcji pyłu zawieszonego $PM_{2,5}$. Autorzy podkreślili konieczność dynamicznej kalibracji oraz kompensacji wpływu wilgotności w czasie rzeczywistym, przeprowadzając jednocześnie analizę długoterminowej stabilności czujników.

W pracy [96] zaprezentowano zastosowanie platform TinyML do predykcji stężeń ozonu w czasie rzeczywistym z wykorzystaniem niskokosztowych urządzeń, takich jak Arduino Nano BLE Sense. System wykorzystuje pomiary temperatury i ciśnienia do estymacji stężeń CO, działa w czasie rzeczywistym i może być montowany na bezzałogowych statkach powietrznych. W badaniu [214] poruszono problem kalibracji niskokosztowych czujników PM, których popularność wynika z niskiej ceny i mobilności. Autorzy zaproponowali wykorzystanie metod uczenia transferowego, umożliwiających efektywną kalibrację przy minimalnym czasie współpracy z urządzeniem referencyjnym. Przedstawiono podejście meta-uczenia, wykorzystujące dane z innych czujników oraz ograniczony zbiór danych z kalibrowanego urządzenia. Wyniki dla frakcji $PM_{2,5}$ wykazały przewagę tej metody nad innymi porównywanymi rozwiązaniami. W pracy [213] zaproponowano tworzenie map stężeń frakcji pyłu zawieszonego $PM_{2,5}$ o wysokiej rozdzielczości poprzez połączenie danych z 320 mobilnych czujników montowanych na pojazdach oraz 52 stacji stacjonarnych. Zastosowano algorytmy uczenia maszynowego do korekcji danych i interpolacji przestrzennej. Uzyskano mapy o rozdzielczości 500 m i odświeżaniu co 5 minut.

W artykule [201] przedstawiono niskokosztowy, wieloparametrowy system monitorowania jakości powietrza, wykorzystujący algorytmy uczenia maszynowego do kalibracji. System umożliwiał wykonanie jednoczesnego pomiaru stężeń pyłów ($PM_{2,5}$, PM_{10}) oraz gazów (SO_2, NO_2, CO, O_3), uwzględniając wpływ temperatury i wilgotności. Opracowany model predykcyjny porównano z danymi referencyjnymi, uzyskując bardzo wysoką dokładność (R^2 do 0,99) przy niskich wartościach RMSE i MAE. W pracy [157] zaprezentowano platformę niskokosztowego monitoringu jakości powietrza opartą na bezzałogowym statku powietrznym wyposażonym w czujniki do pomiaru CO, O_3 oraz NO_2. System zwiększał mobilność i dostępność pomiarów w trudno dostępnych lokalizacjach, eliminując ograniczenia tradycyjnych sieci pomiarowych. Dane były przesyłane w czasie rzeczywistym za pomocą komunikacji LoRa i wizualizowane w aplikacji internetowej. Testy w kontrolowanej przestrzeni powietrznej potwierdziły skuteczność i praktyczność zaproponowanego rozwiązania.

Uczenie maszynowe W kontekście badań związanych z wykorzystaniem metod uczenia maszynowego, w pracy [167] przeanalizowano prognozowanie stężeń dwutlenku azotu (NO_2) na podstawie historycznych danych o zanieczyszczeniach i danych meteorologicznych. Skupiając się na Erfurcie (Niemcy), model uwzględniał zależności czasowe oraz lokalne czynniki, takie jak ruch drogowy czy wydarzenia specjalne. Dodatkowo wykorzystywał sezonowość zanieczyszczeń i siedem parametrów meteorologicznych do prognozowania stężeń NO_2 w nadchodzących godzinach. Wyniki eksperymentów wykazały wyższą dokładność prognoz w porównaniu z innymi modelami, szczególnie w okresach odchyleń od typowych wzorców sezonowych, np. podczas świąt, co może wspierać strategie zarządzania emisjami, w tym regulację ruchu drogowego.

W badaniu [59] zaproponowano model ANN do krótkoterminowego prognozowania emisji CO_2 w sektorze zbożowym Apulii (południowe Włochy), mający na celu ocenę wpływu zrównoważonych praktyk rolniczych na redukcję zanieczyszczeń. Model stanowił elastyczną alternatywę dla klasycznych metod, często zbyt sztywnych do modelowania złożonych zjawisk środowiskowych. Model sieci ANN zaproponowano także w pracy [64] do prognozowania dziennych wartości stężeń pyłów zawieszonych frakcji $PM_{2,5}$ i PM_{10}, z wykorzystaniem danych meteorologicznych.

W pracy [3] zastosowano metodę ANN oraz metodę grupowego przetwarzania danych do prognozowania emisji CO_2 w oparciu o źródła energii. Eksperymenty obejmowały dane z pięciu krajów: Iranu, Kuwejtu, Kataru, Arabii Saudyjskiej oraz Zjednoczonych Emiratów Arabskich. Ocena modelu przy użyciu współczynnika determinacji (R^2) potwierdziła jego skuteczność w przewidywaniu emisji CO_2. Cho i in. zaproponował w badaniu [34] zintegrowany model oparty na ANN do prognozowania parametrów środowiska wewnętrznego w budynku szkolnym, w tym stężeń CO_2, pyłów zawieszonych PM_{10} i $PM_{2,5}$. Model został zoptymalizowany z wykorzystaniem wielokryterialnego algorytmu genetycznego i wyeliminował regresję pozorną poprzez wykluczenie nieistotnych zmiennych wejściowych. Dokładność prognoz oceniano przy użyciu miary RMSE, co wykazało wysoką skuteczność i praktyczną przydatność w kontroli środowiska budynków oraz optymalizacji zużycia energii. Badania [185] oceniają zdolność sztucznej sieci neuronowej do prognozowania godzinnych stężeń zanieczyszczeń powietrza oraz wskaźników jakości powietrza dla Ahvaz (Iran), jednego z najbardziej zanieczyszczonych miast z powodu burz pyłowych. Model wykorzystujący dane meteorologiczne, stężenia zanieczyszczeń oraz czynniki czasowe osiągnął wyższy współczynnik korelacji (R) i niższy RMSE dla sześciu głównych zanieczyszczeń w porównaniu z innymi modelami.

W pracy [123] zastosowano sieci neuronowe w połączeniu z danymi meteorologicznymi do prognozowania godzinnych stężeń pyłu zawieszonego frakcji $PM_{2,5}$ na podstawie danych ze stacji monitoringu Aotizhongxin w Pekinie. Zaproponowana metoda wykazała wysoką dokładność prognoz, osiągając wyższe wartości R^2 w porównaniu z innymi podejściami. Badania [80, 72] dotyczyły prognozowania stężeń pyłu zawieszonego frakcji $PM_{2,5}$ i PM_{10} w Liaocheng i Chongqing (Chiny) z użyciem ANN. Model składał się z 11 parametrów wejściowych, warstwy ukrytej z 6 neuronami i pojedynczego wyjścia. W pracy [71] badano zastosowanie ANN oraz falowych sieci neuronowych (ang. *wavelet transform neural network* – WANN) do prognozowania dziennych stężeń $PM_{2,5}$ w Szanghaju. Modele wykorzystywały zmienne meteorologiczne oraz wcześniejsze stężenia zanieczyszczeń. WANN z funkcją aktywacji *tansig–purelin* przewyższała ANN na wszystkich etapach oceny, wykazując wysoką dokładność prognoz i potwierdzając przydatność do wspierania strategii zarządzania jakością powietrza.

W badaniu [156] zaproponowano model czasoprzestrzennej sieci CNN i LSTM do prognozowania średnich dziennych stężeń $PM_{2,5}$ w Pekinie (Chiny). Model wykorzystał zarówno liniowe, jak i nieliniowe korelacje między parametrami docelowymi a obserwacyjnymi, wykorzystując historyczne dane dotyczące jakości powietrza i meteorologii. Dzięki efektywnej ekstrakcji cech za pomocą sieci neuronowej CNN i odzwierciedlaniu długoterminowych trendów danych za pomocą LSTM, model uzyskał dokładne i szybkie prognozy. W porównaniu z modelami perceptronu wielowarstwowego i LSTM, metoda CNN-LSTM wykazała lepszą stabilność i wydajność w przewidywaniu jakości powietrza. Badanie przedstawione w pracy [180] przedstawiło model głębokiego uczenia, który łączy CNN z LSTM w celu przewidywania godzinnych stężeń całkowitych cząstek zawieszonych (TSP). Sieć CNN pełniła funkcję procesora danych, wykorzystując ekstraktory cech oparte na statystycznie istotnych opóźnionych zmiennych, natomiast sieć LSTM wykorzystała nowy schemat mapowania cech do przewidywania kolejnego godzinnego stężenia TSP. Obszerne testy proponowanego modelu wykazały jego lepszą wydajność w porównaniu z zespołem pięciu innych modeli uczenia maszynowego.

W pracy [25] wykorzystano modele sieci CNN do prognozowania stężeń PM_{10} i $PM_{2,5}$ w Polsce w czasie rzeczywistym. Ocena, oparta na pomiarach R^2 i RMSE, potwierdziła wysoką dokładność modeli w prognozowaniu analizowanych stężeń. W artykule [49] przedstawiono model CNN do prognozowania jakości powietrza w kontekście frakcji $PM_{2,5}$. Proponowany model integruje model CNN w celu wychwycenia lokalnych trendów i korelacji przestrzennych, z dwukierunkową siecią Bi-LSTM w celu ustalenia zależności czasowo-przestrzennych. Obszerne eksperymenty z rzeczywistymi zbiorami danych pokazały, że model osiąga wysoką dokładność prognozowania pyłu zawieszonego $PM_{2,5}$.

Bekkar i in. w pracy [13] rozważali prognozowanie stężeń pyłu zawieszonego w Pekinie (Chiny) z wykorzystaniem metod głębokiego uczenia. W artykule wykorzystano sieci CNN i LSTM do przewidywania godzinnych stężeń pyłu zawieszonego, wykorzystując dane meteorologiczne, a także historyczne dane dotyczące zanieczyszczeń oraz stężenia z pobliskich stacji jako cechy czasoprzestrzenne. Porównano wydajność różnych metod głębokiego uczenia, w tym LSTM, Bi-LSTM, GRU, Bi-GRU i CNN, a model CNN-LSTM przewyższał wszystkie pozostałe pod względem dokładności predykcyjnej. Model łączący splotowe sieci CNN i LSTM do prognozowania zanieczyszczenia powietrza z wykorzystaniem danych z czujników przedstawiono w pracy [63]. Uwzględniono zarówno modele jednowymiarowe, jak i wielowymiarowe. Model przetestowano z wykorzystaniem danych z Barcelony, Kocaeli i Stambułu, wykazując znaczną poprawę dokładności prognozowania w porównaniu z tradycyjnymi modelami LSTM. Zastosowano również transfer uczenia (ang. *transfer learning*).

W pracy [120] opracowano model CNN-LSTM do prognozowania stężenia $PM_{2,5}$ w Pekinie na najbliższe 24 godziny. Model ten łączył zalety sieci CNN i LSTM: CNN efektywnie wyodrębnia cechy związane z jakością powietrza, podczas gdy LSTM przechwytuje długoterminowy kontekst czasowy danych. Model pobierał dane dotyczące jakości powietrza z ostatnich 7 dni i przewidywał stężenie pyłu zawieszonego $PM_{2,5}$ na następny dzień. W badaniu porównano cztery modele: jednowymiarowy LSTM, wielowymiarowy LSTM, jednowymiarowy CNN-LSTM i wielowymiarowy CNN-LSTM. Wyniki wskazują, że proponowany wielowymiarowy model CNN-LSTM przewyższał inne, osiągając najlepszą wydajność przy niskim poziomie błędu i krótkim czasie uczenia.

Autorzy [143] opracowali nowy model prognozowania jakości powietrza oparty na splotowym autoenkoderze Bi-LSTM z mechanizmem uwagi. Badanie przeprowadzono w Busan (Korea Południowa) z wykorzystaniem danych dotyczących stężenia pyłu zawieszonego, warunków meteorologicznych i natężenia ruchu drogowego uzyskanych za pomocą algorytmu YOLO. Model skutecznie uchwycił zależności czasowo-przestrzenne i znacznie przewyższył inne podejścia zarówno w prognozach krótkoterminowych, jak i długoterminowych. Praca [38] zaproponowała nowy model prognozowania jakości powietrza, który łączy autoenkoder wariacyjny i mechanizm uwagi. Metoda została eksperymentalnie zweryfikowana z wykorzystaniem danych o zanieczyszczeniach z czterech stanów USA. Ocena oparta na sześciu metrykach statystycznych wykazała, że model skutecznie prognozował stężenia różnych zanieczyszczeń w wielu lokalizacjach, przewyższając istniejące modele, w tym LSTM, i GRU. W pracy [129] zaproponowano model łączący mechanizm uwagi, sieć LSTM i XGBoost do przewidywania stężeń pyłu zawieszonego frakcji $PM_{2,5}$, uwzględniając wpływ kierunku i prędkości wiatru na transport zanieczyszczeń. Dane wejściowe obejmowały pomiary stężeń pyłu zawieszonego $PM_{2,5}$ z sąsiednich obszarów oraz prognozy meteorologiczne. Eksperymenty przeprowadzono z wykorzystaniem rzeczywistych danych $PM_{2,5}$ i prognoz pogody.

W badaniu [54] zaproponowano model głębokiego uczenia łączący CNN, GRU i mechanizm uwagi w celu prognozowania jakości powietrza. CNN wyodrębniło cechy przestrzenne z obrazów, GRU zarejestrowało zależności czasowe, a mechanizm uwagi dostosował wpływ cech, aby zmniejszyć fluktuacje. Przetestowany na zbiorze danych z Szanghaju model przewyższył inne metody. W [127] zaproponowano zaawansowany model Seq2Seq, który zastąpił sieć RNN mechanizmem uwagi z osadzeniem pozycyjnym. Taka modyfikacja skróciła czas uczenia i zwiększyła odporność modelu. Eksperymenty przeprowadzone w Centrum Olimpijskim i stacjach monitorujących Dongsi w Pekinie potwierdziły, że proponowany model przewyższył istniejące metody, szczególnie w przewidywaniu nagłych zmian wartości stężeń $PM_{2,5}$.

W pracy [1] przeanalizowano skuteczność prognozowania stężeń PM na wschodnim wybrzeżu Półwyspu Malajskiego, gdzie zanieczyszczenie powietrza stanowi istotny problem środowiskowy. Porównano model liniowy oparty na wielokrotnej regresji liniowej z dwoma modelami nieliniowymi: perceptronem wielowarstwowym oraz siecią radialnych funkcji bazowych, wykorzystując historyczne dane meteorologiczne i dane o zanieczyszczeniach z obszarów miejskich, podmiejskich i wiejskich. Najlepsze wyniki uzyskano dla modelu radialnych funkcji bazowych, który zmniejszył błędy prognozowania o 78,9% w obszarach miejskich, 32,1% w podmiejskich oraz 39,8% w wiejskich. Wyniki te podkreślają złożoność relacji pomiędzy stężeniami PM_{10} a czynnikami je warunkującymi oraz wskazują na przewagę modeli nieliniowych, w szczególności sztucznych sieci neuronowych, w zadaniach krótkoterminowego prognozowania. Istotną rolę danych meteorologicznych w modelowaniu jakości powietrza potwierdzono również w pracach [67, 177]. W [67] zaproponowano zintegrowane podejście do prognozowania maksymalnych stężeń ozonu w mikrolokalizacjach miejskich z wykorzystaniem perceptronów wielowarstwowych. Uzyskano poprawę dokładności prognoz jednodniowych, szczególnie na obszarach o niskiej rozdzielczości lokalnej. Z kolei w badaniu [177] wykazano wyższość perceptronu wielowarstwowego nad wielokrotną regresją liniową w prognozowaniu stężeń SO_2 w Teheranie.

Zastosowanie głębokich i rekurencyjnych sieci neuronowych do prognozowania zanieczyszczeń powietrza potwierdzono w kolejnych badaniach. W [110] wykorzystano rekurencyjne sieci neuronowe do estymacji stężeń SO_2 i PM_{10} w Sakarya (Turcja), uzyskując silne korelacje oraz niskie wartości RMSE. Podobnie w [27] zaproponowano zagregowany model LSTM do prognozowania jakości powietrza na podstawie danych z wielu stacji monitoringu na Tajwanie, wykazując jego przewagę nad klasycznymi metodami uczenia maszynowego. Skuteczność sieci radialnych funkcji bazowych potwierdzono również w [186], gdzie zastosowano je do prognozowania jakości powietrza w Zhengzhou i Szanghaju. W artykule [57] rozważano zastosowanie sieci neuronowej autoregresyjnej z rozszerzonym czynnikiem do prognozowania wartości zanieczyszczenia NO_x w Madrycie, wykorzystując opóźnione wartości NO_x, dane meteorologiczne i czynniki ukryte. Model przewyższał konkurencyjne metody, zapewniając dokładniejsze prognozy zanieczyszczenia powietrza. Prognozowanie stężeń NO_x rozważano także w [10].

W [70] porównano standardowe sieci ANN z falkowymi sieciami neuronowymi w prognozowaniu wskaźnika API, wykazując przewagę modeli WANN. Z kolei w [126] zaproponowano ważony model hybrydowy oparty na transformacji falkowej oraz architekturach LSTM, GRU i BiLSTM, który przewyższał pojedyncze modele w prognozowaniu stężeń NO_2 w Pekinie. Dynamiczny rozwój metod opartych na grafowych sieciach neuronowych i mechanizmach uwagi znajduje odzwierciedlenie w licznych pracach. Modele wykorzystujące korelacje czasoprzestrzenne, takie jak GCN-LSTM [139], hierarchiczne architektury DL-Air czy modele fuzji czasowej [82], osiągają wysoką skuteczność w prognozowaniu wielu zanieczyszczeń jednocześnie. Podkreślono to w pracach, m.in. [212, 87, 152, 29, 192, 77], w których uwzględniono zależności przestrzenne pomiędzy stacjami pomiarowymi, wpływ warunków meteorologicznych oraz heterogeniczność systemów monitoringu. Uzupełnieniem badań prognostycznych są systemy działające w czasie rzeczywistym [50].

Kompleksowy przegląd podsumowujący techniki wykorzystujące głębokie uczenie do prognozowania zanieczyszczenia powietrza przedstawiono w pracy [188]. W badaniu podkreślono skuteczność głębokiego uczenia w poprawie dokładności i wiarygodności prognoz jakości powietrza, zwracając uwagę na jego zdolność do modelowania złożonych, nieliniowych zależności między czynnikami meteorologicznymi, stężeniami zanieczyszczeń i innymi zmiennymi wpływającymi. Omówiono również wyzwania związane z wdrażaniem tych technik w rzeczywistych systemach prognozowania jakości powietrza, takie jak jakość danych i koszt obliczeniowy.

Wang i in. w pracy [202] zaproponowali zaawansowany system wczesnego ostrzegania, składający się z dwóch modułów: prognozowania oraz oceny jakości powietrza. Moduł prognostyczny łączył metody wstępnego przetwarzania danych, optymalizacji numerycznej oraz prognozowania interwałowego, umożliwiając uwzględnienie niepewności predykcji. Moduł ewaluacji, oparty na modelu chmury, pozwalał na transformację danych jakościowych w ilościowe miary jakości powietrza. Symulacje przeprowadzone dla danych z miasta Dalian (Chiny) potwierdziły wysoką skuteczność i uniwersalność zaproponowanego rozwiązania. W [228] przedstawiono podejście do prognozowania jakości powietrza oparte na metodach uczenia maszynowego, umożliwiające przewidywanie stężeń zanieczyszczeń w interwałach godzinowych. Zastosowano strategię uczenia wielozadaniowego z uwzględnieniem danych meteorologicznych z wcześniejszych dni. Wyniki eksperymentów wykazały przewagę proponowanego rozwiązania nad klasycznymi modelami regresyjnymi. W badaniu [229] przeanalizowano skuteczność metod uczenia zespołowego w prognozowaniu stężeń $PM_{2,5}$. W modelach uwzględniono zarówno dane klimatyczne, jak i wskaźniki jakości powietrza z miast sąsiednich. Ocenę skuteczności przeprowadzono na zbiorach danych obejmujących pomiary $PM_{2,5}$ z Pekinu, Wuhan oraz Shenzhen, potwierdzając wysoką jakość predykcji.

W pracy [115] zaproponowano model oparty na wzmocnieniu gradientowym do prognozowania stężeń $PM_{2,5}$ na Tajwanie. Analiza porównawcza obejmowała dane z Tajpej oraz Londynu i wykazała zbliżoną skuteczność predykcyjną dla obu miast, co autorzy wiążą z podobnymi uwarunkowaniami topograficznymi. W badaniach [53] opracowano trzy modele uczenia maszynowego: wielokrotną regresję liniową, regresję lasów losowych oraz model addytywny, przeznaczone do prognozowania stężeń pyłów zawieszonych. Wszystkie modele osiągnęły wysoką dokładność, przy czym najlepsze wyniki uzyskano dla modelu addytywnego. Praca przedstawiła również uniwersalną metodykę analizy zmienności PM, możliwą do adaptacji w różnych regionach geograficznych. W [182] zastosowano model Prophet do krótkoterminowego i długoterminowego prognozowania zanieczyszczeń powietrza w Seulu, w tym $PM_{2,5}$ i PM_{10} oraz O_3, NO_2, SO_2 i CO. Model wykazał niższe błędy predykcyjne w porównaniu z alternatywnymi metodami, szczególnie dla pyłów zawieszonych. W badaniu [95] zaproponowano metodę estymacji stężeń benzenu C_6H_6 w obszarach miejskich. Wykorzystano zależności pomiędzy innymi gazami oraz algorytmy uczenia maszynowego, osiągając znaczną poprawę dokładności względem klasycznych metod. Podejście to stanowi efektywną alternatywę dla kosztownego wdrażania dedykowanych czujników. W pracy [136] przeanalizowno poprawę monitorowania jakości powietrza w Chinach poprzez integrację danych satelitarnych o wysokiej rozdzielczości z informacjami meteorologicznymi i chemicznymi z modeli WRF (ang. *Weather Research and Forecasting model*) i CMAQ (ang. *Community Multiscale Air Quality model*). Zastosowano metody asymilacji danych oraz probabilistyczne modele. Uzyskane rezultaty umożliwiają dokładniejsze monitorowanie i prognozowanie długoterminowych zmian stężeń NO i CO. W pracy [134] zaproponowano model uczenia maszynowego integrujący dane z numerycznego modelu WRF-Chem (ang. *Weather Research and Forecasting model with Chemistry*), pomiary naziemne oraz dane meteorologiczne do prognozowania dziennych stężeń $PM_{2,5}$ w Szanghaju. Uzyskano istotną poprawę dokładności w porównaniu z samym modelem WRF-Chem, co potwierdziło zasadność łączenia modeli fizycznych z metodami uczenia maszynowego w prognozowaniu jakości powietrza.

Hybrydowy model PSO-SVM do krótkoterminowego prognozowania stężeń zanieczyszczeń powietrza został zaproponowany w pracy [30]. Metoda łączyła analizę czynników wpływu, klasteryzację danych wejściowych oraz optymalizację parametrów, osiągając lepsze wyniki niż klasyczne modele SVM oraz podejścia oparte na algorytmach genetycznych. Studium przypadku dla Pekinu potwierdziło skuteczność zaproponowanego rozwiązania. W artykule [118] przedstawiono system prognozowania jakości powietrza dla Makau, oparty na pięciu metodach: ANN, RF, XGBoost, SVM oraz MLR. Modele zastosowano do prognozowania stężeń pyłu zawieszonego frakcji $PM_{2,5}$, PM_{10} oraz CO w horyzontach 24 i 48 godzin. Wszystkie algorytmy uzyskały dobre wyniki dla prognoz krótkoterminowych, a najlepszą skuteczność wykazały modele RF i SVM. Prognozowanie wskaźnika jakości powietrza rozważono także w [168].

Metody statystyczne W obszarze wykorzystanych metod statystycznych, badanie przedstawione w pracy [227] skoncentrowano na analizie stężeń $PM_{2,5}$ w Fuzhou (Chiny) z wykorzystaniem modelu ARIMA. Wyniki wskazały sezonowe wahania $PM_{2,5}$, z wyższymi wartościami zimą i niższymi latem. Analiza korelacji Spearmana wykazała dodatnią zależność stężeń $PM_{2,5}$ ze stężeniami PM_{10}, SO_2 i NO_2 oraz ujemną z parametrami meteorologicznymi. W pracy [84] stężenia frakcji PM_{10} w Taiyuan (Chiny) analizowano przy użyciu modelu ARMA/ARIMA opartego na transformacji falkowej. Badania [21] dotyczyły jakości powietrza w Lahore (Pakistan), ujawniając, że stężenia $PM_{2,5}$ i PM_{10} przekraczają krajowe normy środowiskowe. Analiza korelacji wskazała na istotny wpływ ruchu drogowego na wzrost emisji. Prognozowanie przyszłych stężeń $PM_{2,5}$ zrealizowano za pomocą modelu SARIMA. W pracy [6] prognozowano miesięczne stężenia PM_{10} w Erzurum (Turcja). Dane historyczne poddano oczyszczeniu z trendów sezonowych i dekompozycji na trzy poziomy. Modele ARIMA zastosowano do każdej podserii, a model WT-ARIMA przewyższył metodę ARIMA, wykazując wysoką dokładność i przydatność do wczesnego ostrzegania w regionach o dużym zanieczyszczeniu powietrza. Model statystyczny do prognozowania stężeń $PM_{2,5}$ i PM_{10} opracowano także w [2].

W artykule [222] wykorzystano popularne urządzenia LoRa i niskokosztowe czujniki do monitorowania stężeń PM oraz zaproponowano metodę krótkoterminowego prognozowania (do 2 godzin) z użyciem modeli ARIMA i VARMA. W pracy [109] wykorzystano modele ARMA i ARIMA do prognozowania dziennych średnich stężeń O_3, CO, NO i NO_2 w Delhi. Dane poddano transformacji stabilizującej wariancję, a optymalne parametry modelu wybrano na podstawie kryteriów informacyjnych i funkcji autokorelacji. W pracy [60] zastosowano model ARIMA do prognozowania jakości powietrza w prowincji Hunan (Chiny). Model okazał się wiarygodny w przewidywaniu stężeń $PM_{2,5}$, PM_{10}, SO_2 i CO. W badaniu przedstawionym w [8] zastosowano regresję kwantylową do prognozowania stężeń NO_2 w Madrycie.

Metody hybrydowe W literaturze zaproponowano także liczne metody hybrydowe do prognozowania stężeń zanieczyszczeń powietrza, obejmujące zarówno klasyczne modele statystyczne, jak i algorytmy uczenia maszynowego. Wczesne badania koncentrowały się na wykorzystaniu regresji liniowej oraz modeli szeregów czasowych, takich jak AR, ARMA, ARIMA i SARIMA, często łączonych ze sztucznymi sieciami neuronowymi w celu poprawy jakości prognoz [41, 145, 5, 173]. Istotną grupę stanowią modele oparte na sztucznych sieciach neuronowych, w tym perceptronach wielowarstwowych, sieciach RBF, ESN oraz metodach regresji, stosowanych do prognozowania stężeń NO, NO_2 oraz PM_{10} i $PM_{2,5}$ [52, 93, 148, 165, 171].

Badane są także modele łączące algorytmy uczenia maszynowego [48, 116, 174] z technikami dekompozycji sygnału oraz algorytmami optymalizacji. Przykłady takich podejść obejmują wykorzystanie EMD, EEMD i CEEMD w połączeniu z ANN, SVR lub LSTM, optymalizowanych z użyciem PSO, SSA, algorytmów genetycznych, MVO oraz innych metod metaheurystycznych [4, 43, 81, 130, 196, 229]. Przykładowo, w pracach [79, 100] połączono algorytm SVM z algorytmem optymalizacji rojem cząstek do prognozowania stężeń PM_{10}, uzyskując istotną poprawę dokładności względem klasycznych modeli. W ostatnich latach szczególną uwagę poświęcono metodom głębokiego uczenia, zwłaszcza sieciom LSTM oraz ich wariantom, które wykazują wysoką skuteczność w modelowaniu złożonych i nieliniowych zależności czasowych [36, 44, 117, 176, 181, 216]. Kolejnym przykładem może być zastosowanie adaptacyjnego systemu neuro-rozmytego (ANFIS), który w pracy [225] został wykorzystany do prognozowania stężeń CO, SO_2, O_3 oraz NO_2. Podejścia te pozwalają na skuteczniejsze modelowanie niestacjonarnych szeregów czasowych oraz redukcję błędów prognoz. Jednym z aktualnych kierunków jest stosowanie modeli opartych na mechanizmie uwagi wraz z sieciami dwukierunkowymi, np. Bi-LSTM [46].

Coraz częściej stosowane są również metody zespołowe oraz podejścia łączące różne paradygmaty obliczeniowe, w których wyniki wielu modeli są integrowane w celu poprawy jakości prognoz [26, 69, 154, 197]. Zastosowania obejmują zarówno prognozowanie stężeń zanieczyszczeń (PM, NO_2, O_3, SO_2, CO), jak i estymację emisji z wykorzystaniem danych meteorologicznych, środowiskowych oraz społeczno-ekonomicznych [85, 122, 133, 149, 151, 166]. Inne hybrydowe metody prognozowania stężeń analizowano również w pracach [61, 140, 198].

W pracy [209] zbadano właściwości higroskopijne PM, które wpływają na rozpraszanie i absorpcję światła. Analizowano zachowanie pyłu grubego (CPM) i drobnego (frakcja $PM_{2,5}$) w różnych warunkach pogodowych oraz ich wpływ na widoczność. Zastosowano model regresji cenzurowanej do analizy zależności między stężeniami PM a czynnikami meteorologicznymi, umożliwiając obliczenie optycznego współczynnika wzrostu higroskopijnego i wzrostu masy higroskopijnej. Wyniki pokazują, że stężenia CPM i $PM_{2,5}$ negatywnie wpływają na widoczność, a wilgotność względna istotnie modyfikuje te efekty. Wykorzystanie modeli transportu chemicznego WRF-Chem i WRF-CMAQ przeanalizowano w pracach [183].

Z przedstawionego przeglądu literatury wynika, że mimo intensywnego rozwoju metod prognozowania jakości powietrza, w tym podejść opartych na uczeniu maszynowym i głębokim uczeniu, znaczna część badań prowadzona jest w ograniczonych warunkach przestrzennych. Proponowane metody są najczęściej testowane na niewielkiej liczbie stacji pomiarowych, zlokalizowanych w pojedynczych miastach lub regionach o zbliżonych charakterystykach środowiskowych. Taki sposób ewaluacji utrudnia ocenę zdolności modeli do generalizacji oraz ich skuteczności w zróżnicowanych warunkach, obejmujących zarówno obszary silnie zurbanizowane i przemysłowe, jak i regiony o relatywnie niskim poziomie zanieczyszczenia.

1.1.2 Krajowe podejścia do prognozowania stężeń zanieczyszczeń powietrza z wykorzystaniem uczenia maszynowego

W literaturze stosunkowo niewiele jest badań, w których zaprezentowane zostały wyniki analiz w zakresie prognozowania stężeń zanieczyszczeń powietrza w oparciu o dane pochodzące z dużej liczby lokalizacji jednocześnie. Większość dotychczasowych prac koncentruje się na pojedynczych miastach lub wybranych stacjach pomiarowych, co ogranicza możliwość uchwycenia pełnej zmienności przestrzennej i identyfikacji czynników warunkujących powstawanie epizodów zanieczyszczeń. Ograniczony zakres danych może potencjalnie ograniczać skuteczność proponowanych metod, ponieważ nie pozwala w pełni ocenić ich wydajności w prognozowaniu stężeń w różnych warunkach, np. w obszarach o bardzo wysokim stopniu zanieczyszczenia (aglomeracje miejskie) lub o relatywnie niskim zanieczyszczeniu (miejscowości uzdrowiskowe) [102, 107, 224].

Przykładowo, w artykule [161] opisano projekt niskokosztowego systemu pomiaru jakości powietrza montowanego na bezzałogowym statku powietrznym, przeznaczonego do monitoringu powietrza w Opolu. System wykorzystał czujnik PMS7003 do pomiarów $PM_{2,5}$ i PM_{10}, czujniki MQ do detekcji gazów (NO_2, CO) oraz moduł GPS. Wyniki wykazały dobrą korelację danych pozyskanych z UAV z pomiarami naziemnymi. Praca potwierdziła także potencjał dronów jako narzędzi do szybkiej i ekonomicznej identyfikacji obszarów zagrożonych smogiem. Możliwość wykorzystania sieci NARX do prognozowania stężenia PM_{10} pochodzącego z transportu drogowego w Szczecinie zbadano w pracy [99]. Analiza opierała się na danych pomiarowych stężeń PM_{10} oraz parametrach meteorologicznych z okresu 2020–2023, testując różne architektury sieci neuronowych pod kątem dokładności prognozy.

Z kolei w pracy [111] opisano system prognozowania zmian stężenia pyłów frakcji PM_{10} i $PM_{2,5}$ oparty na danych emisyjnych z automatycznych stacji pomiarowych Wojewódzkiego

Inspektoratu Ochrony Środowiska w Warszawie oraz prognozach meteorologicznych z Interdyscyplinarnego Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego. Wykorzystano sztuczne sieci neuronowe i maszyny wektorów nośnych w trybie regresji, a także dekompozycję falkową dla poprawy dokładności prognozy.

W badaniu [224] przeanalizowano szereg czasowy danych dotyczących jakości powietrza w oparciu o sieć 52 niskokosztowych czujników rozmieszczonych w Krakowie i jego okolicach, rejestrujących dane co godzinę. Wyniki pokazały, że sieć niskokosztowych czujników w połączeniu z uczeniem maszynowym skutecznie identyfikuje ekstremalne epizody zanieczyszczeń. W badaniu [102] autorzy zaproponowali metodę krótkoterminowego prognozowania stężeń PM_{10} w Krakowie, wykorzystując dane z sodaru do analizy właściwości fizycznych atmosfery. Badania opierały się na danych historycznych z lat 2017–2018 oraz na niezależnych epizodach wysokiego zanieczyszczenia z października 2021 – marca 2022 w Krakowie. Wyniki wskazały, że metoda z użyciem sodaru jest szybka, tania i obiecująca w przewidywaniu przekroczeń dopuszczalnych stężeń frakcji PM_{10}. W [66] zbadano wpływ różnych czynników na epizody wysokiego stężenia PM_{10} w Krakowie, wykorzystując algorytm RF. Analiza opierała się na danych z publicznego monitoringu jakości powietrza oraz lokalnej stacji meteorologicznej, z zastosowaniem klasyfikacji przypadków, w tym w sytucjach podwyższonych stężeń PM_{10}. Zastosowanie współczesnych sieci neuronowych do prognozowania stężeń frakcji PM_{10} w Krakowie i jego okolicach analizowano także w pracy [101]. W [155] oceniono wpływ asymilacji danych o ozonie powierzchniowym i drobnym aerozolu na dokładność prognoz. Wykorzystano model WRF-Chem oraz narzędzie *Grid-point Statistical Interpolation*, pokazując poprawę dokładności prognoz O_3. Badanie [184] obejmowało analizę stężeń SO_2 w Krakowie. Innym badaniem wykorzystującym stację pomiarową z regionu Małopolski jest praca [65], w której analizie poddano miasto Tarnów.

W badaniu [158] opisano rozwój modeli sztucznych sieci neuronowych do prognozowania maksymalnego godzinowego stężenia ozonu przyziemnego w lokalizacjach miejskich i wiejskich w centralnej Polsce (stacje Belsk i Warszawa). Modele trenowano na danych meteorologicznych i wcześniejszych stężeniach O_3 z okresu kwiecień–wrzesień 2015, a jako zestaw testowy wykorzystano dane z 2014 roku. Wyniki pokazały umiarkowane błędy prognozy oraz tendencję do niewielkiego przeszacowania wartości w Belsku i niedoszacowania w Warszawie, przy czym największe błędy występowały przy nagłych spadkach stężenia ozonu. Badanie [150] wykorzystało sztuczne sieci neuronowe do prognozowania godzinnych stężeń frakcji PM_{10} w Trójmieście w okresach zimowych. Modele uwzględniające dane pomiarowe i czynniki meteorologiczne osiągnęły zadowalające wyniki (R^2 od 0,452 do 0,848), potwierdzając skuteczność ANN w przewidywaniu wartości stężeń zanieczyszczeń. Praca [37] dotyczyła krótkoterminowego prognozowania pyłów zawieszonych PM_{10} i $PM_{2,5}$ w czterech dużych aglomeracjach Polski na podstawie 10-letnich danych pomiarowych. Testowano cztery modele: regresję krokową, lasy losowe, XGBoost i sieci neuronowe. Wyniki podkreślają przydatność algorytmów uczenia maszynowego i znaczenie zmiennych meteorologicznych w modelowaniu jakości powietrza. Prognozowanie godzinowych stężeń pyłów zawieszonych w Pszczynie (woj. śląskie) przedstawiono w [24].

Kujawska i in. porównała w pracy [107] różne modele uczenia maszynowego do prognozowania frakcji PM_{10} w Lublinie, uwzględniając dane meteorologiczne i chemiczne. Sieć neuronowa uzyskała najlepsze wyniki ($R = 0{,}89$), skutecznie przewidując ryzyko przekroczeń norm. W badaniu [106] przeanalizowano prognozy jakości powietrza w Polsce przy użyciu modelu WRF-Chem. W pracy [35] przeanalizowano stężenia PM_{10} w zimowym okresie grudzień–styczeń w latach 2009–2015 w odniesieniu do prędkości wiatru, na podstawie danych z trzech stacji Wojewódzkiego Inspektoratu Ochrony Środowiska w Poznaniu, w tym stacji w Piotrkowicach koło dużej elektrociepłowni. Analiza umożliwiła identyfikację epizodów smogowych (tzw. „czarny smog") oraz weryfikację poprawności pomiarów imisji. Wyniki wykazały, że przy niskich prędkościach wiatru stężenia PM_{10} były wyższe, a zanieczyszczenia mogły przemieszczać się zgodnie z kierunkiem wiatru lub utrzymywać lokalnie przez dłuższy czas. W pracy [58] zaproponowano wykorzystanie perceptronowych modeli regresyjnych do prognozowania stężeń zanieczyszczeń powietrza. Badania przeprowadzono na podstawie danych historycznych z Zabrza oraz Złotego Potoku (województwo śląskie).

Nielicznymi pracami, w których zaprezentowano modele prognostyczne zwalidowane na większej liczbie stacji pomiarowych, są artykuły [83, 88, 89, 169]. Prognozowanie stężeń trzech frakcji pyłu zawieszonego (PM_1, $PM_{2,5}$ oraz PM_{10}) zrealizowano w badaniu [88]. Użyto kilku modeli uczenia maszynowego, w którym porównano efektywność algorytmów RF i DT z sieciami RNN oraz LSTM. Jakość prognoz oceniono na podstawie danych historycznych, pochodzących ze stacji pomiarowych zlokalizowanych w miastach: Bytom, Kraków i Żywiec. W pracy [83] oceniono dokładność prognozowania stężeń $PM_{2,5}$ w oparciu o dane stężeń PM_{10} z wykorzystaniem sieci MLP w kilku monitoringowych stacjach w Polsce. Analiza opierała się na długoterminowych, godzinnych danych pomiarowych, przy uwzględnieniu silnej korelacji między stężeniami $PM_{2,5}$ i PM_{10}. Wyniki pokazały, że modele MLP mogą skutecznie estymować stężenia frakcji $PM_{2,5}$ nawet w takich miejscach, gdzie bezpośrednie pomiary tej frakcji nie są prowadzone. Eksperymenty przeprowadzono na bazie danych z Olsztyna, Płocka, Radomia, Lublina, Kędzierzyna-Koźla i Jeleniej Góry.

Natomiast w badaniu [89] analizowano długoterminowe trendy i zmienność przestrzenną stężeń PM_{10} w Polsce w latach 2019–2024, wykorzystując dane z Europejskiego Systemu Monitoringu Jakości Powietrza dla dziewięciu lokalizacji obejmujących aglomeracje (Warszawa, Kraków, Katowice), miasta (Wrocław, Poznań) oraz uzdrowiska (Ciechocinek, Lądek-Zdrój, Świnoujście). Zastosowano zintegrowane podejście statystyczne, łączące liniowe modele mieszane (ang. *linear mixed models* – LMM), modele addytywne i regresję kwantylową. Wyniki pokazały, że niezależnie od różnorodności lokalizacji, zastosowane metody statystyczne dobrze uchwyciły zarówno trendy czasowe, jak i sezonowe zmienności stężeń PM_{10}, co wskazuje na ich wysoką skuteczność w analizie zróżnicowanych warunków środowiskowych. Badanie [169] analizowało sezonowe zmiany stężenia i składu chemicznego $PM_{2,5}$ w Katowicach, Gdańsku i Diablej Górze. Najwyższe stężenia odnotowano zimą w Katowicach, gdzie odnotowano podwyższone stężenia m.in. aerozoli wtórnych, i benzo(a)pirenu (BaP).

1.1.3 Podsumowanie

Analiza zaprezentowanej literatury wskazuje, że prognozowanie stężeń zanieczyszczeń powietrza stanowi złożone zadanie badawcze, realizowane z wykorzystaniem różnych grup metod. W prowadzonych badaniach zastosowanie znajdują zarówno klasyczne podejścia statystyczne, jak i metody uczenia maszynowego, uczenia głębokiego oraz rozwiązania hybrydowe.

W literaturze światowej obserwuje się stopniowe przechodzenie od prostszych modeli liniowych i autoregresyjnych w kierunku metod umożliwiających odwzorowanie nieliniowych zależności występujących w danych środowiskowych. Szczególne znaczenie mają w tym zakresie metody uczenia maszynowego oraz uczenia głębokiego, które pozwalają na analizę dużych zbiorów danych, uwzględnianie wielu zmiennych wejściowych oraz modelowanie zależności czasowych i czasowo-przestrzennych. W wielu badaniach wykorzystywane są również rozwiązania hybrydowe, łączące różne algorytmy w celu poprawy dokładności prognoz oraz ograniczenia błędów predykcji.

Istotnym elementem rozwoju metod prognostycznych jest także rozszerzanie zakresu danych wejściowych. Oprócz historycznych stężeń zanieczyszczeń coraz częściej wykorzystywane są dane meteorologiczne, informacje sezonowe, dane pochodzące z sąsiednich stacji pomiarowych, a także wyniki pomiarów z niskokosztowych czujników, sieci sensorowych, systemów IoT oraz platform mobilnych. Takie podejście umożliwia lepsze odwzorowanie zmienności jakości powietrza, jednak jednocześnie zwiększa wymagania dotyczące jakości danych, ich kompletności, wstępnego przetwarzania oraz walidacji modeli.

Analiza badań prowadzonych w Polsce wskazuje, że uwzględnione w przeglądzie krajowe prace koncentrują się przede wszystkim na prognozowaniu stężeń pyłów zawieszonych frakcji PM_{10} i $PM_{2,5}$, z wykorzystaniem zarówno klasycznych metod statystycznych, jak i algorytmów uczenia maszynowego oraz głębokiego. Jednocześnie znaczna część analiz dotyczy wybranych miast, regionów lub ograniczonej liczby stacji pomiarowych. Powoduje to, że ocena skuteczności modeli w zróżnicowanych warunkach środowiskowych, obejmujących odmienne uwarunkowania meteorologiczne, emisyjne i przestrzenne, pozostaje zagadnieniem wymagającym dalszej analizy.

Przeprowadzona analiza literatury w zakresie przedmiotu pracy pozwala stwierdzić, że skuteczność prognozowania stężeń zanieczyszczeń powietrza zależy nie tylko od zastosowanego algorytmu, lecz także od sposobu przygotowania danych, doboru zmiennych wejściowych, przyjętej procedury walidacyjnej oraz charakterystyki analizowanego obszaru. Szczególne znaczenie ma możliwość porównania działania modeli w różnych lokalizacjach i warunkach środowiskowych, ponieważ pozwala to ocenić ich stabilność oraz potencjalną użyteczność praktyczną. Wnioski te stanowiły podstawę do podjęcia badań przedstawionych w niniejszej pracy, których założenia, tezę oraz cel omówiono w kolejnym podrozdziale.

1.2 Motywacja, teza i cel pracy

Analiza literatury wskazuje, że większość badań nad prognozowaniem stężeń zanieczyszczeń powietrza w Polsce ogranicza się do stosunkowo niewielkiej liczby stacji pomiarowych lub regionów o homogenicznych warunkach środowiskowych, co utrudnia ocenę zdolności modeli do generalizacji. Badania wykazują także silne uzależnienie wyników prognoz od lokalnych uwarunkowań geograficznych, klimatycznych, czy charakterystyki źródeł emisji. W konsekwencji, konieczne wydaje się dostosowanie architektury modeli do specyfiki lokalnej, uwzględniając zmienność meteorologiczną i geograficzną oraz stosowanie szerokich zestawów danych obejmujących różne typy lokalizacji. Podejście takie jest kluczowe przy porównywaniu wyników badań oraz wdrażaniu systemów prognozujących stężenia zanieczyszczeń w różnych częściach kraju oraz wskazuje na potrzebę przeprowadzenia analiz w szerszej skali przestrzennej, z uwzględnieniem danych pochodzących z różnorodnych lokalizacji.

W związku z tym zaproponowano następującą **tezę** niniejszej pracy:

Metody głębokiego uczenia, zastosowane do prognozowania stężeń wybranych zanieczyszczeń powietrza (w tym pyłów zawieszonych frakcji PM_{10} i $PM_{2,5}$), na podstawie danych historycznych pochodzących z wielu geograficznie i środowiskowo zróżnicowanych lokalizacji w Polsce, zapewniają wyższą oraz bardziej stabilną jakość predykcji w krótkoterminowych horyzontach prognozy niż klasyczne modele statystyczne i regresyjne w analizowanym zbiorze danych.

Głównym celem rozprawy jest opracowanie, implementacja oraz ocena efektywności nowoczesnych metod głębokiego uczenia w krótkoterminowym prognozowaniu stężeń pyłów zawieszonych i wybranych gazów zanieczyszczających powietrze w Polsce, ze szczególnym uwzględnieniem zróżnicowanych warunków geograficznych, klimatycznych oraz typów obszarów, obejmujących obszary metropolitalne, miasta średniej wielkości i obszary wiejskie.

Istotnym i nowatorskim elementem podejścia jest wykorzystanie danych pochodzących z wielu geograficznie i środowiskowo zróżnicowanych lokalizacji w Polsce, co stanowi główny aspekt nowości i pozwala na analizę zdolności modeli do generalizacji. Skuteczność wybranych architektur głębokiego uczenia porównano z klasycznymi modelami statystycznymi i regresyjnymi, traktowanymi jako metody referencyjne. Oceny dokonano dla wybranych zanieczyszczeń powietrza oraz różnych krótkoterminowych horyzontów prognozy.

Zakres rozprawy obejmuje:

(1) Przegląd literatury dotyczący metod prognozowania stężeń zanieczyszczeń powietrza w Polsce i na świecie, ze szczególnym uwzględnieniem metod opartych na danych, w tym metod statystycznych, metod uczenia maszynowego, uczenia głębokiego oraz rozwiązań hybrydowych;

(2) Pozyskanie, uporządkowanie i przygotowanie historycznych danych pomiarowych pochodzących z automatycznych stacji monitoringu jakości powietrza funkcjonujących w ramach Państwowego Monitoringu Środowiska, udostępnianych przez Główny Inspektorat Ochrony Środowiska, obejmujących:

   (a) Dane godzinowe z dnia 2 stycznia 2022 r. dla ośmiu zanieczyszczeń powietrza: C_6H_6, CO, NO, NO_2, NO_x, O_3, PM_{10} i SO_2, ze stacji zlokalizowanych w Jeleniej Górze, Wrocławiu, Gorzowie Wielkopolskim, Sulęcinie, Zakopanem, Gdańsku, Częstochowie i Koszalinie;

   (b) Dane dobowe z okresu od 1 stycznia do 31 grudnia 2022 r. dla pyłu zawieszonego frakcji PM_{10} oraz As, Cd, Ni i Pb, ze stacji zlokalizowanych w Wałbrzychu, Toruniu, Żaganiu, Krakowie, Mielcu, Katowicach, Puszczy Boreckiej i Szczecinie;

   (c) Dane dotyczące stężeń PM_{10} i $PM_{2,5}$ z okresu od 1 stycznia do 31 grudnia 2022 r. ze stacji w Zielonce Bory Tucholskie, Zakopanem, Częstochowie, Gliwicach/Knurowie oraz Szczecinie, uzupełnione w wybranych wariantach o dane meteorologiczne z platformy Meteostat;

   (d) Dodatkowe dane pomiarowe z lat 2023–2024, wykorzystane do oceny stabilności oraz zdolności uogólniania opracowanych modeli w kolejnych okresach pomiarowych.

(3) Analizę danych pomiarowych pochodzących ze stacji reprezentujących różne warunki środowiskowe i emisyjne w Polsce, w tym lokalizacje miejskie, przemysłowe, górskie, wiejskie oraz obszary tła regionalnego;

(4) Uwzględnienie w badaniach różnych grup zanieczyszczeń powietrza, obejmujących pyły zawieszone frakcji PM_{10} i $PM_{2,5}$, wybrane gazy zanieczyszczające powietrze, takie jak C_6H_6, CO, NO, NO_2, NO_x, O_3 i SO_2, a także As, Cd, Ni i Pb;

(5) Przygotowanie wariantów eksperymentalnych opartych na historycznych stężeniach zanieczyszczeń oraz, w wybranych pracach, na dodatkowych danych meteorologicznych, obejmujących temperaturę, prędkość i kierunek wiatru oraz ciśnienie atmosferyczne;

(6) Opracowanie i implementację metod krótkoterminowego prognozowania stężeń zanieczyszczeń powietrza opartych na architekturach uczenia głębokiego, w tym (m.in.): TCN, FCN, GRU, xLSTM, KAN, TFT oraz modelach autoenkoderowych i hybrydowych;

(7) Przeprowadzenie eksperymentów dla różnych konfiguracji danych wejściowych, w tym z wykorzystaniem okien historycznych obejmujących 3, 5 oraz 7 wcześniejszych wartości pomiarowych, a także wariantów z uwzględnieniem i bez uwzględnienia danych meteorologicznych;

(8) Porównanie skuteczności zaproponowanych metod z metodami referencyjnymi, obejmującymi klasyczne i powszechnie stosowane podejścia prognostyczne, takie jak wygładzanie wykładnicze, metoda Holta-Wintersa, ARIMA, ARIMAX, RF, MLP, RNN, LSTM, k-NNR oraz regresja liniowa;

(9) Ocenę efektywności zaproponowanych metod z wykorzystaniem standardowych miar jakości predykcji, takich jak MAE, MAPE, MSE, RMSE oraz R^2, a także analizę stabilności wyników w odniesieniu do rodzaju zanieczyszczenia, lokalizacji stacji pomiarowej, konfiguracji danych wejściowych oraz analizowanego okresu pomiarowego;

(10) Weryfikację możliwości zastosowania opracowanych metod do krótkoterminowego prognozowania stężeń zanieczyszczeń powietrza w zróżnicowanych warunkach środowiskowych Polski, z uwzględnieniem lokalizacji miejskich, przemysłowych, górskich, wiejskich oraz obszarów tła regionalnego.

Rdzeń metodologiczny rozprawy stanowi zatem analiza i porównanie wybranych architektur głębokiego uczenia w zadaniu prognozowania stężeń zanieczyszczeń powietrza na podstawie danych historycznych, z uwzględnieniem danych pochodzących z wielu lokalizacji. Poszczególne części pracy oraz powiązane publikacje przedstawiają natomiast różne aspekty aplikacyjne tego podejścia, w tym analizę wpływu zróżnicowania przestrzennego danych oraz zdolności modeli do generalizacji.

Rozprawa jest oparta na wynikach badań przedstawionych w publikacjach naukowych tworzących spójny cykl tematyczny, obejmujących artykuły opublikowane oraz artykuł konferencyjny, który został zaprezentowany podczas międzynarodowej konferencji *25th International Conference on Artificial Intelligence and Soft Computing (ICAISC 2026)* i zaakceptowany do publikacji w recenzowanych materiałach Springer LNAI (Lecture Notes in Artificial Intelligence).

1.3 Układ pracy

Praca składa się z pięciu rozdziałów, a jej struktura została dostosowana do formatu rozprawy doktorskiej przygotowanej na podstawie publikacji naukowych powiązanych tematycznie z jej problematyką. W rozdziale 2 formalnie zdefiniowano zadanie prognozowania dla szeregów czasowych, przedstawiono zarys architektur sieci neuronowych stosowanych w rozprawie, scharakteryzowano wykorzystane miary oceny jakości predykcji oraz szczegółowo opisano procedurę walidacyjną zapewniającą poprawność i porównywalność przeprowadzonych eksperymentów. W rozdziale 3 przedstawiono krótką charakterystykę wyników badań i analiz zaprezentowanych szczegółowo w siedmiu pracach naukowych powiązanych tematycznie z rozprawą, w tym pięciu o zasięgu międzynarodowym, indeksowanych w bazie Web of Science, których dane bibliograficzne zestawiono w tab. 1.2. Przegląd literatury został przedstawiony w pracy A1 [18], natomiast prace A2–A7 prezentują wyniki przeprowadzonych analiz i obliczeń [14, 15, 16, 17, 19, 20]. Rozdział 4 przedstawia dodatkowe wyniki badań, którym poddano zaproponowane metody w celu weryfikacji ich skuteczności dla danych pomiarowych pochodzących z lat 2023 i 2024. W rozdziale 5 podsumowano wnioski wynikające z przeprowadzonych badań oraz wskazano możliwe kierunki dalszych analiz. W rozdziale dodatkowym A zamieszczono teksty publikacji, potwierdzenia dotyczące statusu publikacyjnego oraz wymagane oświadczenia, w tym charakterystykę zakresu prac wykonanych przez autora niniejszej rozprawy. Ponadto, w rozdziale oznaczonym jako dodatek B do niniejszej pracy krótko scharakteryzowano poprzednią rozprawę doktorską autora, obronioną w dyscyplinie informatyka techniczna i telekomunikacja. Przedstawiono założenia badawcze, tezę i cel pracy, publikacje oraz podsumowano uzyskane wyniki.

1.4 Konwencje

W niektórych przypadkach, aby uniknąć niejasności, przytoczone są oryginalne (angielskie) nazwy pojęć, których tłumaczenie nie jest jednoznaczne lub mogłoby się okazać kontrowersyjne. W opisach formalnych czcionką pogrubioną oznaczono wektory.

Tabela 1.2. Wykaz artykułów oraz publikacji konferencyjnej przyjętej do druku, powiązanych tematycznie z rozprawą doktorską. W tabeli podano status publikacyjny, punktację według właściwego wykazu ministerialnego oraz wartość pięcioletniego współczynnika Impact Factor (IF), jeżeli dotyczy.

L.p.	**Pełny opis bibliograficzny**	**Pkt.**	**IF**
A1	J. Bernacki, R. Scherer (2025): A Comprehensive Review of Data-Driven Techniques for Air Pollution Concentration Forecasting. Sensors (MDPI); 25(19), 6044, Q2	100	4,0
A2	J. Bernacki (2025): Forecasting the air pollution concentration with neural networks. Urban Climate (Elsevier); Vol. 59 102262, Q1	100	6,9
A3	J. Bernacki (2025): Forecasting the concentration of the components of the particulate matter in Poland using neural networks. Environmental Science and Pollution Research (Springer); Vol. 32, pp 9179–9212, Q1	100	5,8
A4	J. Bernacki (2025): A hybrid transformer and symbolic regression model for weather-dependent particulate matter forecasting in air quality management. Theoretical and Applied Climatology (Springer); Vol. 156, No. 662, Q2	70	3,2
A5	J. Bernacki (2025): Particulate Matter Forecasting Using Hybrid Autoencoders: The Role of Meteorological Data. Annales UMCS Sectio B – Geographia, Geologia, Mineralogia et Petrographia, Vol. 80, pp 265-283, Q3	40	–
A6	J. Bernacki, I. Sówka: Forecasting Particulate Matter with Neural Networks: Accuracy and Explainability using Meteorological Data. Zaakceptowana i przedstawiona podczas 25th International Conference on Artificial Intelligence and Soft Computing	20	–
A7	J. Bernacki, I. Sówka: Forecasting Particulate Matter Concentrations with Deep Neural Networks Using Meteorological Data. Zeszyty Naukowe SGSP 2026, No. 98(1)	70	–
	Σ	**500**	**19,9**

Ze względu na wymogi poprawy dostępności cyfrowej (obejmującej m.in. zalecenia wynikające ze stosowania standardu WCAG 2.2), niniejsza rozprawa została zredagowana z wykorzystaniem czcionki bezszeryfowej.

Rozdział 2

Podstawy metodyczne prognozowania stężeń zanieczyszczeń powietrza z wykorzystaniem uczenia głębokiego

W rozdziale przedstawiono podstawy metodyczne wykorzystane w badaniach dotyczących krótkoterminowego prognozowania stężeń zanieczyszczeń powietrza. W odróżnieniu od przeglądu literatury zaprezentowanego w rozdziale 1, niniejsza część koncentruje się na zagadnieniach bezpośrednio związanych z realizacją analiz własnych, obejmujących modelowanie szeregów czasowych, zastosowanie metod uczenia głębokiego, procedurę walidacyjną, dobór miar jakości predykcji oraz analizę statystyczną wyników. Przedstawione zagadnienia stanowią podstawę metodyczną dla opisu publikacji zaprezentowanych w rozdz. 3 i załączonych w rozdziale dodatkowym A oraz dodatkowych wyników badań przedstawionych w rozdz. 4.

2.1 Szeregi czasowe i zadanie prognozowania

Szeregiem czasowym określa się uporządkowaną sekwencję obserwacji pewnej wielkości w kolejnych punktach czasowych. W niniejszej rozprawie przez szereg czasowy będziemy rozumieć wszystkie obserwacje stężeń wybranych zanieczyszczeń powietrza i towarzyszących im parametrów meteorologicznych, zebrane w kolejnych punktach czasowych na poszczególnych stacjach pomiarowych.

Formalnie, szereg czasowy $\mathbf{y}$ można zdefiniować następująco. Niech:

- $\{\mathbf{y}\}_{t=1}^{T}$ będzie obserwowanym szeregiem czasowym (zawierającym historyczne wartości stężenia zanieczyszczeń powietrza w kolejnych odstępach czasowych);

- t – indeks czasowy (np. godzina, doba);
- T – ostatni znany moment obserwacji.

W praktyce, szereg czasowy może więc być wektorem o następującej postaci:

$$\mathbf{y} = [y_1, y_2, \dots, y_T]$$

gdzie y_T oznacza wartość obserwacji w jednostce czasu t. Zarówno historyczne stężenia zanieczyszczeń (np. C_6H_6, NO_2, $PM_{2,5}$, itd.) , jak i historyczne wartości danych meteorologicznych, powinny być reprezentowane za pomocą osobnych szeregów czasowych [23, 55, 62].

Zadanie prognozowania h-krokowego (ang. *h-step ahead forecasting*) polega na estymacji przyszłych wartości [75, 78]:

$$\hat{y}_{T+h|T} \quad \text{dla} \quad h = 1, 2, \dots, H$$

gdzie:

- $\hat{y}_{T+h|T}$ – prognoza wartości y_{T+h} zbudowana na podstawie informacji dostępnych do momentu T włącznie;
- H – horyzont prognozowania (liczba kroków do przodu, które mają być prognozowane).

2.2 Uczenie głębokie i sztuczne sieci neuronowe

Uczenie głębokie (ang. *deep learning* – DL) jest poddziedziną uczenia maszynowego, w której modele oparte na wielowarstwowych sieciach neuronowych uczą się złożonych reprezentacji danych poprzez automatyczne wydobywanie cech na kolejnych poziomach abstrakcji. Metody uczenia głębokiego wykorzystuje się do wykonywania zadań takich jak klasyfikacja, regresja, prognozowanie czy uczenie reprezentacji. Inspiracją dla uczenia głębokiego są biologiczne sieci neuronowe w mózgu człowieka, w których miliardy połączonych neuronów przetwarzają informacje równocześnie, tworząc złożone wzorce zachowań i percepcji [113]. Głównym narzędziem uczenia głębokiego są sieci neuronowe, które dzięki różnorodnym typom i architekturom mogą być dopasowane do różnych zastosowań, w tym prognozowania danych sekwencyjnych [46, 124, 220].

Sztuczne sieci neuronowe (ang. *artificial neural networks* – ANN) stanowią klasę modeli obliczeniowych inspirowanych sposobem przetwarzania informacji w biologicznych układach nerwowych. Ich istotą jest zdolność do aproksymacji złożonych, nieliniowych zależności pomiędzy danymi wejściowymi i wyjściowymi na podstawie przykładów uczących. Podstawowym elementem sieci neuronowej jest sztuczny neuron, który realizuje operację ważonej sumy sygnałów wejściowych, do której następnie stosowana jest funkcja aktywacji. Formalnie, wyjście pojedynczego neuronu y można zapisać jako

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right), \tag{2.1}$$

gdzie:

- x_i – sygnały wejściowe oraz w_i – odpowiadające im wagi;
- b – wyraz wolny (ang. *bias*);
- $f(\cdot)$ – funkcja aktywacji.

W praktyce najczęściej stosowane są nieliniowe funkcje aktywacji, takie jak funkcja ReLU (ang. *rectified linear unit*), funkcja sigmoidalna czy tangens hiperboliczny.

Neurony organizowane są w warstwy, tworząc architekturę sieci neuronowej. Typowa sieć składa się z warstwy wejściowej, jednej lub wielu warstw ukrytych oraz warstwy wyjściowej. Sieci zawierające więcej niż jedną warstwę ukrytą określane są mianem sieci głębokich (ang. *deep neural networks*), a ich zastosowanie pozwala na hierarchiczne przetwarzanie informacji i ekstrakcję cech o rosnącym poziomie abstrakcji [189].

Proces uczenia sieci neuronowej polega na optymalizacji wartości wag i biasów w celu minimalizacji funkcji kosztu, opisującej różnicę pomiędzy wartościami przewidywanymi przez model a danymi referencyjnymi. W zadaniach regresyjnych, takich jak prognozowanie, powszechnie stosowane są funkcje kosztu oparte na błędach średniokwadratowych lub bezwzględnych. Optymalizacja realizowana jest najczęściej z wykorzystaniem algorytmu wstecznej propagacji błędu (ang. *backpropagation*) w połączeniu z metodami gradientowymi, takimi jak algorytm Adam [170, 211].

Niepożądanym zjawiskiem towarzyszącym uczeniu sztucznych sieci neuronowych jest ryzyko przeuczenia modelu (ang. *overfitting*), polegające na nadmiernym dopasowaniu do danych uczących kosztem zdolności generalizacji. W celu ograniczenia tego zjawiska powszechnie stosowane są techniki regularyzacji, takie jak wczesne zatrzymanie procesu uczenia (ang. *early stopping*), normalizacja danych wejściowych oraz losowe wyłączanie neuronów w trakcie uczenia (ang. *dropout*). Mechanizm ten polega na tymczasowym usuwaniu części połączeń sieciowych, co zapobiega nadmiernej współzależności neuronów i sprzyja uczeniu bardziej uogólnionych reprezentacji. Odpowiedni dobór architektury sieci oraz strategii regularyzacji stanowi istotny element projektowania modeli neuronowych o wysokiej skuteczności predykcyjnej [47, 103].

W kontekście analizy szeregów czasowych rozumianej jako prognozowanie, kluczową rolę odgrywa sposób konstrukcji danych wejściowych. Dane pomiarowe organizowane są zazwyczaj w postaci sekwencji obejmujących określone okno czasowe, na podstawie którego model generuje prognozę dla wybranego horyzontu czasowego.

2.3 Rodzaje sieci neuronowych

Sieci konwolucyjne i temporalne Sieci konwolucyjne (ang. *convolutional neural networks* – CNN) stanowią klasę modeli głębokiego uczenia, w których podstawową operacją przetwarzania danych jest splot. Choć architektury te zostały pierwotnie zaprojektowane z myślą o analizie danych obrazowych, ich koncepcja została z powodzeniem zaadaptowana również do analizy szeregów czasowych, w których dane mają postać jednowymiarowych sekwencji obserwacji uporządkowanych w czasie [121, 153].

W przypadku szeregów czasowych stosuje się najczęściej splot jednowymiarowy, w którym filtr konwolucyjny przesuwany jest wzdłuż osi czasu. Filtr taki stanowi wektor wag o określonej długości, zwanej rozmiarem jądra, i umożliwia lokalne przetwarzanie fragmentów sekwencji wejściowej. Wynikiem operacji splotu jest nowa sekwencja, określana jako mapa cech, odzwierciedlająca obecność określonych wzorców czasowych w danych. Istotnym pojęciem w sieciach konwolucyjnych jest pole recepcyjne (częściej określane jako okno), które określa zakres danych wejściowych wpływających na pojedynczy element mapy cech. Zasięg ten może być dodatkowo kontrolowany poprzez stosowanie parametrów takich jak krok przesuwania filtra (ang. *stride*) oraz dopełnianie sekwencji wejściowej (ang. *padding*). W architekturach konwolucyjnych często stosuje się również operacje redukcji wymiarowości, takie jak warstwy grupujące (ang. *pooling*), które agregują informacje z lokalnych fragmentów map cech [147].

Rozszerzeniem klasycznego splotu są sploty z dylatacją (ang. *dilations*), które są stosowane w sieciach konwolucyjnych temporalnych (TCN). Celem splotów z dylatacją jest wprowadzenie pomiędzy kolejne elementy filtra „przerw" [112, 219]. Pozwala to na zwiększenie pola recepcyjnego bez proporcjonalnego wzrostu liczby parametrów modelu. Mechanizm ten znajduje zastosowanie w analizie długich sekwencji czasowych, gdzie istotne są zależności występujące w różnych skalach czasowych [97, 153].

Sieci konwolucyjne mogą być stosowane zarówno do analizy pojedynczych szeregów czasowych, jak i do przetwarzania danych wielowymiarowych, w których każdy kanał wejściowy reprezentuje odrębną zmienną czasową. Dzięki temu architektury konwolucyjne stanowią uniwersalne narzędzie do ekstrakcji cech z danych sekwencyjnych, niezależnie od dziedziny zastosowania [56, 194, 203].

Sieci rekurencyjne Sieci rekurencyjne (RNN) stanowią jedną z klasycznych grup modeli neuronowych przeznaczonych do analizy danych sekwencyjnych, w tym szeregów czasowych. Ich kluczową cechą jest obecność połączeń zwrotnych, umożliwiających przekazywanie informacji pomiędzy kolejnymi krokami czasowymi. Dzięki temu model może wykorzystywać informację o przeszłych obserwacjach przy generowaniu predykcji dla kolejnych chwil czasowych. Klasyczne sieci RNN cechują się jednak ograniczoną zdolnością do modelowania długoterminowych zależności czasowych, co wynika z problemu zanikania gradientu podczas procesu uczenia [42].

W celu przezwyciężenia w literaturze pojawiły się bardziej zaawansowane architektury rekurencyjne, takie jak sieci LSTM [68] oraz GRU [40], wykorzystujące mechanizmy bramkowania do kontrolowania przepływu informacji w czasie. Rozwiązania te umożliwiają selektywne zapamiętywanie i „zapominanie" informacji, co znacząco poprawia skuteczność modelowania zależności długookresowych w szeregach czasowych. W ostatnich latach pojawiły się również rozszerzenia klasycznych architektur, takie jak xLSTM, które wprowadzają dodatkowe mechanizmy adaptacyjne oraz skalowalne struktury pamięci, zwiększając stabilność uczenia i zdolność modelowania złożonych dynamik czasowych w długich sekwencjach [12].

Autoenkodery Autoenkodery (ang. *autoencoders*) są architekturą sieci neuronowej, pierwotnie stosowaną do kompresji i rekonstrukcji danych. W kontekście prognozowania szeregów czasowych ich architektura może zostać przystosowana do generowania wartości przyszłych obserwacji zamiast wiernej rekonstrukcji wejścia. Standardowy autoenkoder składa się z *enkodera*, który przekształca fragment historycznego szeregu czasowego do zwartej reprezentacji (tzw. ukrytej) oraz z *dekodera*, który generuje prognozy kolejnych punktów w szeregu czasowym [11, 160].

Prognozujące autoenkodery wykorzystują zdolność enkodera do wychwycenia kluczowych wzorców i zależności w danych historycznych, w tym sezonowości, trendów oraz krótkoterminowych fluktuacji. Dekoder natomiast przekształca tę reprezentację w prognozę przyszłych wartości. W praktyce stosuje się różne warianty autoenkoderów, m.in. konwolucyjno-rekurencyjne (CAE) [73, 217], które w enkoderze wykorzystują warstwy konwolucyjne oraz komórki LSTM lub GRU, autoenkodery hierarchiczne (HA) [206] oraz autoenkodery wariancyjne (VAE) [45], które wprowadzają probabilistyczną reprezentację ukrytą, pozwalającą uwzględnić niepewność prognoz.

Sieci typu transformer Modele typu transformer stanowią klasę architektur opartych na mechanizmie uwagi (ang. *attention*), pierwotnie zaprojektowanych do przetwarzania sekwencji w zadaniach przetwarzania języka naturalnego. W ostatnich latach zyskały one jednak istotne znaczenie również w analizie i prognozowaniu szeregów czasowych, stanowiąc alternatywę dla klasycznych sieci rekurencyjnych oraz konwolucyjnych.

Podstawowym elementem transformera jest mechanizm samo-uwagi (ang. *self-attention*), który umożliwia modelowi jednoczesne uwzględnienie zależności pomiędzy wszystkimi elementami sekwencji wejściowej, niezależnie od ich odległości czasowej. W odróżnieniu od sieci rekurencyjnych, transformery nie przetwarzają danych sekwencyjnie, lecz równolegle, co pozwala znacząco przyspieszyć proces uczenia. Zależności temporalne są kodowane za pomocą mechanizmów pozycjonowania (ang. *positional encoding*), które dostarczają informacji o kolejności obserwacji w szeregu czasowym [125, 195].

W kontekście prognozowania szeregów czasowych, modele transformerowe są wykorzystywane zarówno do predykcji jednego kroku w przód, jak i prognoz wielohoryzontowych. Dane wejściowe mogą obejmować wyłącznie historyczne obserwacje analizowanej zmiennej lub być rozszerzone o dodatkowe cechy. Architektura transformera umożliwia elastyczne łączenie informacji pochodzących z różnych źródeł bez konieczności ręcznego definiowania struktury zależności czasowych [205, 226].

2.4 Procedura walidacyjna i schemat eksperymentalny

Procedura walidacyjna została zaprojektowana w sposób zapewniający porównywalność modeli, zachowanie porządku czasowego danych oraz eliminację ryzyka przecieku informacji między etapem uczenia i testowania modeli. W szczególności, w analizach wykorzystywano podejście oparte na przesuwającym się oknie czasowym (ang. *sliding window*), co pozwalało na realistyczne odwzorowanie warunków prognozowania szeregów czasowych. W wybranych eksperymentach stosowano również chronologiczny podział danych na zbiory uczący, walidacyjny i testowy.

Proces doboru architektur oraz ich parametrów realizowano eksperymentalnie, poprzez analizę jakości uzyskiwanych predykcji dla różnych konfiguracji modeli. Niezależnie od szczegółowych ustawień poszczególnych eksperymentów, zachowano spójne założenia metodologiczne, obejmujące wykorzystanie wyłącznie danych historycznych na etapie uczenia, stosowanie wspólnych schematów oceny oraz zapewnienie porównywalności analizowanych podejść w ramach danego eksperymentu. Dodatkowo, w celu ograniczenia ryzyka przeuczenia modeli, w wybranych pracach zastosowano mechanizm wczesnego zatrzymania (ang. *early stopping*), polegający na przerwaniu procesu uczenia w przypadku braku poprawy jakości predykcji w kolejnych epokach (z cierpliwością wynoszącą 20 epok).

2.5 Miary błędów i ocena jakości predykcji

Ocena jakości modeli prognostycznych w zadaniach prognozowania pozwala porównywać określone metody, a także określić ich jakość. Niech x_i oznacza wartość obserwowaną, a y_i wartość prognozowaną. W niniejszej pracy wykorzystano klasyczne miary błędów regresji, do których należą:

- Średni błąd bezwzględny (ang. *mean absolute error* – MAE) [207]:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - x_i|, \tag{2.2}$$

- Średni procentowy błąd bezwzględny (ang. *mean absolute percentage error* – MAPE) [33]:

$$\mathrm{MAPE} = \frac{100}{N}\sum_{i=1}^{N} \left|\frac{x_i - y_i}{x_i}\right|. \tag{2.3}$$

- Błąd średniokwadratowy (ang. *mean squared error* – MSE) [33]:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N} (y_i - x_i)^2, \tag{2.4}$$

- Pierwiastek błędu średniokwadratowego (ang. *root mean squared error* – RMSE) [207]:

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}. \tag{2.5}$$

- Współczynnik determinacji (R^2) [33]:

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{N}(x_i - y_i)^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}, \tag{2.6}$$

  gdzie $\bar{x}$ oznacza średnią wartość obserwowaną.

Miary MAE oraz MAPE informują o średniej wielkości błędu niezależnie od kierunku odchylenia. MSE i RMSE są wrażliwe na duże odchylenia, co umożliwia wychwycenie błędów prognozy. Współczynnik R^2 przyjmuje wartości z przedziału $[0, 1]$, gdzie 0 oznacza brak dopasowania, a 1 pełne dopasowanie.

Tabela 2.1. Miary błędów stosowane w cyklu publikacji tworzącym rozprawę

Miara	Artykuły
MAE	A2, A3, A4, A5, A6, A7
MAPE	A2
MSE	A2, A3, A4, A5, A6
RMSE	A2, A3, A4, A6
R^2	A6, A7

Wykorzystanie określonych miar w rozważanych publikacjach zestawiono w tab. 2.1.

Miary oceny jakości modeli prognostycznych z publikacji A4–A7, zostały obliczone dla wartości znormalizowanych. W przeprowadzonych eksperymentach nie wykonywano operacji odwrotnej transformacji (odnormalizowania) wartości prognozowanych ani obserwowanych przed wyznaczeniem wskaźników błędu. Dzięki temu możliwe było bezpośrednie porównanie jakości modeli trenowanych i ocenianych przy wykorzystaniu znormalizowanych danych. Wyjątek stanowią publikacje A2 i A3, w których nie stosowano normalizacji danych, a miary jakości wyznaczono bezpośrednio dla wartości w ich oryginalnej skali.

2.6 Analiza statystyczna wyników

Weryfikacja statystyczna oparta na testowaniu hipotez pozwala odpowiedzieć na pytanie, czy prognozy wygenerowane przez proponowane modele, różnią się istotnie od danych historycznych. W niniejszej rozprawie wszystkie testy przeprowadzono na poziomie istotności $\alpha = 0{,}05$.

Testy normalności rozkładu

Weryfikację hipotez rozpoczyna się od testów normalności rozkładu. W tym celu można użyć testu Shapiro-Wilka, a hipotezy zdefiniowane są następująco [179]:

- $\mathcal{H}_0$: Analizowane próby są zgodne z rozkładem normalnym;
- $\mathcal{H}_1$: Analizowane próby nie są zgodne z rozkładem normalnym.

W celu weryfikacji hipotezy zerowej oblicza się statystykę testu Shapiro–Wilka S, a następnie odpowiadającą jej wartość p (ang. *p-value*). Otrzymaną wartość p porównuje się z przyjętym poziomem istotności α. Jeżeli $p > \alpha$, nie ma podstaw do odrzucenia hipotezy zerowej $\mathcal{H}_0$. Natomiast, gdy $p \leqslant \alpha$, hipoteza zerowa zostaje odrzucona na rzecz hipotezy alternatywnej

$\mathcal{H}_1$. Równoważnie, jeśli wartość statystyki S jest większa od wartości krytycznej testu, nie ma podstaw do odrzucenia hipotezy zerowej [178].

ANOVA Kruskala-Wallisa

Kolejnym etapem analizy jest ocena istotności różnic pomiędzy badanymi próbami. W przypadku spełnienia założeń dotyczących normalności rozkładu danych, stosuje się parametryczną analizę wariancji (ANOVA), umożliwiającą weryfikację istotności różnic średnich pomiędzy wieloma grupami. W sytuacji, gdy założenia te nie są spełnione, wykorzystuje się nieparametryczną analizę wariancji ANOVA Kruskala–Wallisa (zwaną również jednokierunkową analizą wariancji dla rang), stanowiącą rozszerzenie testu U Manna–Whitneya na przypadek więcej niż dwóch populacji. Test ten służy do weryfikacji hipotezy o braku istotnych różnic pomiędzy medianami w $k > 2$ próbach [104, 105]. Hipotezy dla k grup zdefiniowane są następująco:

- $\mathcal{H}_0$: $\theta_1 = \theta_2 = \cdots = \theta_k$ (wszystkie mediany są równe);
- $\mathcal{H}_1$: Nie wszystkie mediany θ_j są równe ($j = 1, 2, \ldots, k$).

W celu podjęcia decyzji o odrzuceniu lub braku podstaw do odrzucenia hipotezy zerowej $\mathcal{H}_0$, oblicza się wartość statystyki testowej H, a następnie odpowiadającą jej wartość p. Wartość p porównuje się z przyjętym poziomem istotności α. Jeżeli $p > \alpha$, nie ma podstaw do odrzucenia hipotezy zerowej. W przeciwnym przypadku, gdy $p \leqslant \alpha$, hipotezę zerową odrzuca się na rzecz hipotezy alternatywnej $\mathcal{H}_1$. Równoważnie, decyzję można podjąć poprzez porównanie wartości statystyki testowej H z wartością krytyczną. Jeżeli wartość statystyki testowej przekracza wartość krytyczną, hipotezę zerową $\mathcal{H}_0$ należy odrzucić [141].

2.7 Podsumowanie

W rozdziale przedstawiono podstawy metodyczne wykorzystane w badaniach dotyczących krótkoterminowego prognozowania stężeń zanieczyszczeń powietrza. Omówiono zagadnienie szeregów czasowych oraz specyfikę zadania prognozowania, wskazując na znaczenie danych historycznych w modelowaniu zmian stężeń zanieczyszczeń powietrza. Następnie scharakteryzowano wybrane metody uczenia głębokiego oraz rodzaje sieci neuronowych, które stanowią podstawę opracowanych i analizowanych w pracy rozwiązań prognostycznych.

Istotnym elementem rozdziału było także przedstawienie procedury walidacyjnej i schematu eksperymentalnego, umożliwiających ocenę skuteczności proponowanych metod w sposób uporządkowany i porównywalny. Uwzględnienie standardowych miar błędów predykcji, takich jak

MAE, MSE, RMSE, MAPE oraz współczynnik R^2, pozwala na ilościową ocenę jakości prognoz oraz porównanie wyników uzyskanych dla różnych modeli, lokalizacji, okresów pomiarowych i konfiguracji danych wejściowych.

Przedstawione zagadnienia porządkują aparat metodyczny zastosowany w dalszej części pracy. Mają one bezpośrednie znaczenie dla interpretacji wyników uzyskanych w pracach naukowych powiązanych tematycznie z rozprawą, ponieważ pozwalają na spójne odniesienie zastosowanych architektur modeli, procedur walidacyjnych oraz miar oceny jakości predykcji do poszczególnych etapów przeprowadzonych badań.

Rozdział 3

Charakterystyka wyników badań i analiz zaprezentowanych w pracach naukowych A1–A7

W niniejszym rozdziale przedstawiono charakterystykę prac naukowych A1–A7 powiązanych tematycznie z rozprawą doktorską, obejmujących zarówno przegląd aktualnego stanu wiedzy, jak i wyniki przeprowadzonych analiz oraz badań wspierających realizację celu pracy i weryfikację postawionej tezy badawczej. Tytuły kolejnych podrozdziałów zostały sformułowane tak, aby odzwierciedlały główny zakres merytoryczny oraz zasadniczy wkład poszczególnych prac, tj. przegląd metod *data-driven* w prognozowaniu jakości powietrza, zastosowanie sieci neuronowych i metod uczenia głębokiego, wykorzystanie modeli hybrydowych, transformerowych i autoenkoderów, a także znaczenie danych meteorologicznych oraz wyjaśnialności modeli prognostycznych. W każdym podrozdziale przedstawiono krótkie streszczenie omawianej pracy, obejmujące podstawowe założenia badawcze, zastosowaną metodykę oraz najważniejsze uzyskane wyniki. W pracach A2–A7 wykorzystano historyczne dane pomiarowe ze stacji monitoringu jakości powietrza, udostępnione przez GIOŚ w ramach Państwowego Monitoringu Środowiska. Dane meteorologiczne wykorzystane w pracach A4–A7 pobrano z platformy Meteostat [144].

3.1 Stan wiedzy w zakresie metod *data-driven* stosowanych do prognozowania stężeń zanieczyszczeń powietrza (artykuł A1)

W pracy A1 [18] dokonano kompleksowego przeglądu metod prognozowania stężeń zanieczyszczeń powietrza opartych na metodach wykorzystujących dane pomiarowe, który obejmował zarówno klasyczne techniki statystyczne (m.in. ARIMA, SARIMA, regresja wieloraka), metody uczenia maszynowego (m.in. SVM, RF, XGBoost), modele głębokiego uczenia (m.in. LSTM, GRU, CNN, sieci typu *transformer*, autoenkodery, GCN), hybrydowe (m.in. CNN+LSTM, ARIMA+ANN), jak i zaawansowane technologie sensoryczne (m.in. sensory mobilne, IoT, LoRaWAN, drony, TinyML). Przeanalizowano aktualny stan badań nad prognozowaniem kluczowych zanieczyszczeń, takich jak (m.in.) pył zawieszony, tlenki azotu, ozon, tlenki węgla oraz tlenki siarki. W pracy A1 wskazano, że modele głębokiego uczenia, zwłaszcza hybrydowe oraz oparte na mechanizmach uwagi, są najbardziej obiecujące ze względu na lepszą zdolność do uchwycenia nieliniowych zależności czasowo-przestrzennych, przy jednoczesnej przewadze nad tradycyjnymi podejściami statystycznymi w zakresie dokładności prognoz. W artykule przedyskutowano także istotne otwarte problemy, takie jak jakość danych, interpretowalność modeli oraz integracja heterogenicznych systemów sensorycznych.

Na podstawie przeprowadzonej analizy literatury można stwierdzić, że prognozowanie stężeń zanieczyszczeń powietrza coraz częściej opiera się na metodach uczenia głębokiego, które dzięki zdolności modelowania złożonych zależności czasowo-przestrzennych osiągają wyższą skuteczność niż klasyczne modele statystyczne. Jednocześnie rozwój technologii sensorycznych umożliwia pozyskiwanie danych o coraz większej rozdzielczości przestrzennej i czasowej, co stwarza nowe możliwości budowy dokładniejszych modeli prognostycznych. Istotnym nadal wyzwaniem pozostaje zapewnienie wysokiej jakości danych, zwiększenie interpretowalności modeli oraz efektywna integracja danych pochodzących z heterogenicznych źródeł.

W tabeli 3.1 podsumowano wyniki przeglądu literatury opisanego w pracy A1.

Tabela 3.1. Zestawienie wybranych metod prognozowania stężeń zanieczyszczeń powietrza opracowane na podstawie przeglądu literatury przedstawionego w artykule A1

Grupa metod	**Przykładowe metody**	**Przykładowe źródła literaturowe**	**Główne wnioski**	**Znaczenie dla kolejnych prezentowanych prac**
Techniki sensoryczne	Czujniki mobilne, sieci IoT, LoRaWAN, czujniki niskokosztowe, drony, TinyML	[7, 90, 96, 108, 157, 161, 213, 214, 223]	Rozwiązania sensoryczne zwiększają rozdzielczość czasową i przestrzenną danych, jednak wymagają kalibracji, kontroli jakości oraz integracji z danymi referencyjnymi.	Wskazanie znaczenia jakości, kompletności i wiarygodności danych wejściowych wykorzystywanych w modelowaniu prognostycznym.
Metody statystyczne	AR, ARMA, ARIMA, SARIMA, regresja wieloraka, regresja kwantylowa	[2, 6, 8, 21, 60, 84, 109, 222, 227]	Metody statystyczne są interpretowalne i użyteczne jako punkt odniesienia, jednak ich skuteczność może być ograniczona w przypadku danych nieliniowych, niestacjonarnych i silnie zmiennych sezonowo.	Uzasadnienie wykorzystania metod klasycznych jako metod referencyjnych w ocenie skuteczności proponowanych modeli głębokiego uczenia.

Tabela 3.1. Syntetyczne zestawienie (kontynuacja)

Grupa metod	**Przykładowe metody**	**Przykładowe źródła literaturowe**	**Główne wnioski**	**Znaczenie dla kolejnych prezentowanych prac**
Metody uczenia maszynowego	RF, SVM, XGBoost, GAM, Prophet, drzewa decyzyjne, metody zespołowe	[3, 34, 59, 64, 71, 80, 123, 167, 185]	Metody uczenia maszynowego umożliwiają odwzorowanie bardziej złożonych zależności niż klasyczne modele statystyczne, lecz wymagają odpowiedniego doboru zmiennych wejściowych i procedur walidacyjnych.	Uzasadnienie potrzeby porównania metod głębokiego uczenia z wybranymi algorytmami uczenia maszynowego oraz zastosowania uporządkowanej procedury oceny jakości predykcji.
Metody uczenia głębokiego	ANN, RNN, LSTM, GRU, Bi-LSTM, CNN, GCN, autoenkodery, sieci typu *transformer*	[13, 25, 38, 49, 63, 120, 143, 156, 180]	Modele głębokiego uczenia są szczególnie przydatne w modelowaniu złożonych zależności czasowych i czasowo-przestrzennych oraz w analizie dużych zbiorów danych środowiskowych.	Bezpośrednie uzasadnienie zastosowania architektur głębokiego uczenia w kolejnych pracach badawczych wchodzących w skład cyklu.

Tabela 3.1. Syntetyczne zestawienie (kontynuacja)

Grupa metod	**Przykładowe metody**	**Przykładowe źródła literaturowe**	**Główne wnioski**	**Znaczenie dla kolejnych prezentowanych prac**
Modele hybrydowe	ARIMA+ANN, CNN+LSTM, GCN+LSTM, PSO+LSTM, SSA+LSTM, modele z mechanizmami uwagi	[5, 41, 44, 107, 117, 145, 148, 176, 197]	Modele hybrydowe łączą zalety różnych podejść i często prowadzą do poprawy dokładności prognoz, jednak zwiększają złożoność obliczeniową oraz wymagania interpretacyjne.	Uzasadnienie rozwijania i testowania bardziej złożonych architektur, w tym modeli autoenkoderowych, transformerowych oraz rozwiązań wykorzystujących dodatkowe dane meteorologiczne.
Problemy otwarte i kierunki rozwoju	Jakość danych, braki danych, interpretowalność modeli, heterogeniczność źródeł danych, systemy wczesnego ostrzegania	[61, 102, 116, 151, 166, 171, 174, 209, 216]	Pomimo dużej liczby badań istotne pozostają problemy związane z odpornością modeli, ich skalowalnością, wyjaśnialnością oraz możliwością stosowania w różnych warunkach środowiskowych.	Uzasadnienie oceny stabilności modeli, analizy wpływu danych meteorologicznych oraz weryfikacji metod na danych pochodzących z różnych lokalizacji i okresów pomiarowych.

3.2 Zastosowanie sieci neuronowych do prognozowania stężeń wybranych zanieczyszczeń powietrza (artykuł A2)

Celem prac zaprezentowanych w publikacji A2 było sprawdzenie, czy nowoczesne modele głębokiego uczenia mogą skutecznie prognozować krótkoterminowe stężenia zanieczyszczeń powietrza na podstawie historycznych danych pomiarowych. Dodatkowo, praca miała na celu porównanie skuteczności zaproponowanych architektur z klasycznymi metodami prognostycznymi oraz ocenę wpływu długości okna wejściowego na jakość predykcji. Istotnym elementem celu badawczego było również sprawdzenie działania modeli nie dla pojedynczej stacji pomiarowej, lecz w układzie obejmującym kilka lokalizacji i kilka zanieczyszczeń, co zwiększa znaczenie środowiskowe i aplikacyjne uzyskanych wyników.

W badaniach wykorzystano rzeczywiste dane pochodzące z automatycznych stacji monitoringu jakości powietrza (działających w ramach PMŚ). Analiza obejmowała osiem miast (Jelenią Górę, Wrocław, Gorzów Wielkopolski, Sulęcin, Zakopane, Gdańsk, Częstochowę i Koszalin) oraz osiem analizowanych zanieczyszczeń powietrza (C_6H_6, CO, NO, NO_2, NO_x, O_3, PM_{10} i SO_2). Uwzględnienie różnych lokalizacji pozwoliło ocenić skuteczność modeli w odmiennych warunkach przestrzennych i emisyjnych, natomiast analiza kilku zanieczyszczeń umożliwiła sprawdzenie, czy proponowane podejście zachowuje skuteczność dla danych o różnej dynamice czasowej. Szczególne znaczenie ma również fakt, że badania odnosiły się do sezonu grzewczego, który w warunkach polskich jest okresem kluczowym z punktu widzenia występowania podwyższonych stężeń zanieczyszczeń powietrza.

W publikacji zaproponowano i przetestowano cztery architektury głębokiego uczenia: temporalną sieć konwolucyjną (TCN), sieć Kołmogorowa–Arnolda (KAN), sieć w pełni konwolucyjną (FCN) oraz rekurencyjną sieć GRU. Modele analizowano dla różnych długości okna wejściowego, obejmujących 3, 5 oraz 7 wartości historycznych. Pozwoliło to ocenić, w jakim stopniu zakres informacji historycznej wpływa na jakość krótkoterminowego prognozowania stężeń zanieczyszczeń powietrza. Zastosowanie kilku architektur i kilku wariantów okna wejściowego umożliwiło systematyczne porównanie skuteczności modeli w zależności od lokalizacji, rodzaju zanieczyszczenia oraz konfiguracji eksperymentalnej. Proponowane metody przetestowano w oparciu o dane z roku 2022 na bazie stacji pomiarowych działających w ramach programu PMŚ.

Skuteczność proponowanych modeli porównano z metodami referencyjnymi, obejmującymi klasyczne podejścia prognostyczne i regresyjne, takie jak wygładzanie wykładnicze, metoda Holta-Wintersa, ARIMA oraz algorytm RF. Takie porównanie było istotne, ponieważ pozwoliło ocenić, czy zastosowanie bardziej złożonych architektur głębokiego uczenia jest uzasadnione poprawą jakości predykcji względem prostszych i dobrze znanych metod. Do oceny jakości prognoz wykorzystano standardowe miary błędu, takie jak MAE, MAPE, MSE i RMSE.

Wyniki uzyskane w publikacji A2 wskazują, że proponowane modele głębokiego uczenia osiągały niższe wartości miar błędu niż metody referencyjne. Przykładowo, tabela 3.2 przedstawia wartości błędu MAE dla stacji Gdańsk Leczkowska (oznaczenie PMŚ: PmGdaLeczkow).

Wybrane wyniki z tej tabeli można zestawić następująco:

Tabela 3.2. Wartości MAE dla stacji pomiarowej PmGdaLeczkow (średnia$^{(1)}$ oznacza średnią uzyskaną dla wszystkich zanieczyszczeń powietrza dla wybranej metody; średnia$^{(2)}$ oznacza średnią zarówno dla proponowanych metod, jak i dla metod bazowych); wyniki na podstawie tabeli A.22 z publikacji A2 [15]

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	średnia$^{(1)}$	średnia$^{(2)}$
min.	0,23	0,42	0,48	3,97	4,79	20,93	4,02	2,16	—	—
maks.	1,24	0,56	2,93	25,81	30,30	66,57	15,45	5,22	—	—
TCN 3	0,08	0,03	0,23	1,81	2,50	6,42	2,28	0,34	1,71	0,09
TCN 5	0,08	0,03	0,16	2,84	3,25	9,79	2,22	0,54	2,36	
TCN 7	0,10	0,04	0,17	2,45	2,66	6,91	2,12	0,33	1,85	
KAN 3	0,06	0,03	0,04	0,37	0,64	7,43	1,74	0,25	1,32	
KAN 5	0,06	0,03	0,08	1,40	1,72	7,75	1,81	0,26	1,64	
KAN 7	0,06	0,03	0,01	0,60	0,74	7,42	1,65	0,24	1,34	
FCN 3	0,00	0,02	0,00	0,35	0,50	7,51	0,27	0,17	1,10	
FCN 5	0,06	0,02	0,05	0,43	0,60	7,88	1,45	0,26	1,34	
FCN 7	0,00	0,02	0,00	0,24	0,50	7,57	0,34	0,23	1,11	
GRU 3	0,00	0,02	0,01	0,18	0,24	7,48	1,32	0,22	1,18	
GRU 5	0,04	0,02	0,02	0,14	0,67	7,80	1,76	0,19	1,33	
GRU 7	0,00	0,03	0,01	0,57	0,24	7,46	0,67	0,12	1,14	
wygł. wykł.	0,15	0,02	0,30	3,51	4,12	9,99	1,80	0,74	2,58	0,15
Holt-Winters	0,13	0,01	0,30	3,90	3,73	10,04	1,45	0,69	2,53	
ARIMA	0,09	0,01	0,19	2,59	3,59	11,55	0,89	0,56	2,43	
RF	0,13	0,01	0,20	2,85	3,72	9,76	1,32	0,68	2,33	

Na podstawie tego zestawienia można wskazać, że dla analizowanej stacji PmGdaLeczkow najniższe średnie wartości MAE uzyskały modele FCN oraz GRU z oknem wejściowym 7. W przypadkach tych średni błąd był wyraźnie niższy niż dla metod referencyjnych, takich jak ARIMA, algorytm RF czy wygładzanie wykładnicze. Szczególnie widoczna jest różnica dla C_6H_6 i CO, dla których modele głębokiego uczenia uzyskały niższe wartości błędu niż klasyczne metody prognostyczne.

Rozszerzenie badań z pracy A2 na lata 2023 i 2024 przedstawiono w rozdz. 4 w tab. 4.1. W tabeli tej pokazano, że dla artykułu A2, po uśrednieniu wyników względem stacji pomiarowych i zanieczyszczeń, proponowane metody również osiągały niższe wartości błędów niż metody referencyjne. Przykładowo, dla danych godzinowych z 2022 r. wartość MAE wyniosła 0,11 dla metod proponowanych oraz 0,19 dla metod referencyjnych, natomiast RMSE odpowiednio 0,14 oraz 0,28. Podobna relacja utrzymywała się również dla danych z lat 2023 i 2024.

Uzyskane wyniki potwierdzają, że w analizowanym zbiorze danych zastosowanie architektur głębokiego uczenia pozwoliło na poprawę jakości krótkoterminowej predykcji względem metod referencyjnych. Istotne jest przy tym, że poprawa ta była widoczna zarówno w przykładzie szczegółowym dla konkretnej stacji pomiarowej, jak i w wynikach uśrednionych. Oznacza to, że skuteczność zaproponowanych modeli nie ograniczała się wyłącznie do pojedynczego przypadku, lecz była obserwowana również w szerszym zestawieniu wyników.

Wyniki publikacji A2 bezpośrednio wspierają główną tezę rozprawy, ponieważ pokazują, że wybrane architektury głębokiego uczenia mogą być skutecznym narzędziem krótkoterminowego prognozowania stężeń zanieczyszczeń powietrza w zróżnicowanych warunkach środowiskowych. Wkład omawianej pracy polega przede wszystkim na eksperymentalnym potwierdzeniu, że modele typu *data-driven* mogą osiągać lepszą jakość predykcji niż metody referencyjne, a także na pokazaniu, że ocena skuteczności takich modeli powinna uwzględniać lokalizację stacji, rodzaj zanieczyszczenia, sezonowość oraz długość okna wejściowego.

Ze względu na to, że ocena modeli została w określonym układzie danych, lokalizacji, substancji i konfiguracji eksperymentalnych, wyniki należy interpretować w odniesieniu do przyjętego zbioru danych oraz zastosowanego schematu walidacji (który został szczegółowo opisany w rozdziale. 2). Naturalnym kierunkiem dalszych badań jest sprawdzenie stabilności zaproponowanych modeli na danych z kolejnych lat, rozszerzenie analiz o dodatkowe zmienne meteorologiczne oraz dalsza ocena zdolności modeli do generalizacji w innych warunkach przestrzennych i czasowych. W tym kontekście publikacja A2 tworzy punkt wyjścia dla kolejnych prac, w których rozwijane są aspekty walidacji, odporności modeli oraz wpływu dodatkowych danych wejściowych na jakość prognoz.

3.3 Prognozowanie stężeń pyłu zawieszonego z wykorzystaniem metod uczenia głębokiego (artykuł A3)

Celem badania przedstawionego w artykule A3 było sprawdzenie jakości prognozowania krótkoterminowych stężeń zanieczyszczeń powietrza nowoczesnych modeli uczenia głębokiego na podstawie historycznych danych pomiarowych. W pracy porównano także skuteczność zaproponowanych architektur z klasycznymi metodami prognostycznymi oraz oceniono wpływ długości okna wejściowego na jakość predykcji. W tym celu wykorzystano rzeczywiste dane dotyczące stężenia pyłu zawieszonego frakcji PM_{10} oraz As, Cd, Ni i Pb, pochodzące z automatycznych stacji monitoringu jakości powietrza, obejmujących osiem lokalizacji: Wałbrzych, Toruń, Żagań, Kraków, Mielec, Katowice, Puszczę Borecką i Szczecin. Uwzględnienie różnych lokalizacji pozwoliło ocenić skuteczność modeli w odmiennych warunkach przestrzennych i emisyjnych, natomiast analiza zanieczyszczeń umożliwiła sprawdzenie, czy proponowane podejście zachowuje skuteczność dla danych o różnej dynamice czasowej. Szczególne znaczenie ma również fakt, że badania objęły sezon grzewczy. Można więc przyjąć, że praca A3 jest rozszerzeniem badań przedstawionych w artykule A2.

W badaniu zaproponowano i przetestowano cztery architektury głębokiego uczenia: rozszerzoną sieć z długą pamięcią krótkotrwałą (xLSTM), sieć Kołmogorowa–Arnolda (KAN), temporalną sieć konwolucyjną (TCN) oraz autoenkoder wariacyjny (VAE). Modele analizowano dla różnych długości okna wejściowego, obejmujących 3, 5 oraz 7 wartości historycznych. Pozwoliło to ocenić, w jakim stopniu zakres informacji historycznej wpływa na jakość krótkoterminowego prognozowania stężeń zanieczyszczeń powietrza. Zastosowanie kilku architektur i kilku wariantów okna wejściowego umożliwiło systematyczne porównanie skuteczności modeli w zależności od lokalizacji, rodzaju zanieczyszczenia oraz konfiguracji eksperymentalnej. Opracowane metody przebadano na danych z roku 2022 ze stacji pomiarowych prowadzonych w ramach programu PMŚ.

Skuteczność proponowanych modeli porównano z metodami referencyjnymi, obejmującymi klasyczne podejścia prognostyczne i regresyjne, takie jak wygładzanie wykładnicze, metoda Holta-Wintersa, ARIMA oraz algorytm RF. Porównanie to pozwoliło ocenić, czy zastosowanie bardziej złożonych architektur głębokiego uczenia jest uzasadnione poprawą jakości predykcji względem prostszych i powszechnie znanych metod. Do oceny jakości prognoz wykorzystano standardowe miary błędu, jak MAE, MSE i RMSE.

Uzyskane w artykule A3 wyniki wskazują, że proponowane modele głębokiego uczenia osiągały niższe wartości miar błędu niż metody referencyjne. Przykładowo, tabela 3.3 przedstawia wartości błędu RMSE dla stacji Toruń Dziewulskiego (oznaczenie PMŚ: KpToruDziewu).

Wyniki wskazują, że proponowane metody uzyskują nieco niższe wartości błędu RMSE niż rozwiązania referencyjne. Najniższe średnie wartości RMSE uzyskano dla modeli KAN oraz TCN z rozmiarem okna 5, dla których średnia wartość błędu wyniosła 2,88. Dla metod bazowych najlepsze wyniki osiągnął model Holt–Winters ze średnim RMSE równym 3,34. Widoczna jest również zależność od długości okna wejściowego, ponieważ dla wszystkich analizowanych architektur najkorzystniejsze wyniki uzyskano dla pięciu próbek wejściowych. Średnia wartość RMSE dla wszystkich proponowanych metod wyniosła 3,28, podczas gdy dla metod referencyjnych osiągnęła wartość 3,49, co wskazuje na przewagę opracowanych rozwiązań.

Rozszerzenie badań z pracy A3 na lata 2023 i 2024 przedstawiono w rozdz. 4 w tab. 4.2. Wykazano, że dla artykułu A3, po uśrednieniu wyników względem stacji pomiarowych i zanieczyszczeń, proponowane metody również osiągały niższe wartości błędów niż metody re-

Tabela 3.3. Wartości RMSE dla stacji KpToruDziewu (średnia$^{(1)}$ oznacza średnią uzyskaną dla wszystkich zanieczyszczeń powietrza dla wybranej metody; średnia$^{(2)}$ oznacza średnią zarówno dla proponowanych metod, jak i dla metod bazowych); wyniki na podstawie tabeli 27 z publikacji A3 [16]

	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}	średnia^1	średnia^2
min.	0,50	0,10	0,50	0,00	2,00	—	—
maks.	2,87	0,42	7,21	0,02	90,76	—	—
xLSTM 3	0,37	0,07	1,42	0,00	15,19	3,41	3,28
xLSTM 5	0,28	0,05	1,05	0,00	15,00	3,28	
xLSTM 7	0,36	0,10	1,78	0,00	14,91	3,43	
KAN 3	0,37	0,07	1,51	0,00	14,13	3,22	
KAN 5	0,27	0,05	1,07	0,00	13,00	2,88	
KAN 7	0,37	0,09	1,85	0,00	14,24	3,31	
TCN 3	0,38	0,07	1,51	0,00	15,00	3,39	
TCN 5	0,27	0,05	1,07	0,00	12,98	2,88	
TCN 7	0,37	0,09	1,85	0,00	14,81	3,43	
VAE	0,38	0,07	1,52	0,00	14,94	3,38	
wygł. wykł.	0,71	0,17	1,70	0,00	14,43	3,40	3,49
Holt-Winters	0,65	0,04	0,73	0,00	15,30	3,34	
ARIMA	0,59	0,14	1,73	0,00	15,78	3,65	
RF	0,53	0,15	1,71	0,00	15,43	3,57	

ferencyjne. Przykładowo, dla danych godzinowych z 2022 r. wartość MSE wyniosła 0,02 dla metod proponowanych oraz 0,08 dla metod referencyjnych, natomiast dla lat 2023 i 2024, odpowiednio: 0,03 i 0,11; 0,03 i 0,12. Analogiczne wnioski płyną z analizy pozostałych miar błędów.

Uzyskane wyniki wskazują, że zastosowanie architektur głębokiego uczenia pozwoliło w analizowanym zbiorze danych na poprawę jakości krótkoterminowej predykcji względem metod referencyjnych. Poprawa ta została wykazana zarówno w przykładzie szczegółowym dla określonej stacji pomiarowej, jak i w wynikach uśrednionych. Prowadzi to do wniosku, że skuteczność proponowanych modeli była obserwowana również w szerszym zestawieniu wyników.

Podobnie, jak w przypadku pracy A2, wyniki zawarte w publikacji A3 bezpośrednio wspierają tezę rozprawy, ponieważ pokazują, że wybrane architektury głębokiego uczenia mogą być skutecznym narzędziem krótkoterminowego prognozowania stężeń zanieczyszczeń powietrza w zróżnicowanych warunkach środowiskowych.

Przeprowadzenie oceny modeli w określonym układzie danych, lokalizacji oraz konfiguracji eksperymentalnej może być jednak pewnym ograniczeniem pracy A3. Wyniki powinny być interpretowane w odniesieniu do wykorzystanego zbioru danych oraz zastosowanego schematu walidacji. Kierunkiem dalszych badań jest sprawdzenie stabilności zaproponowanych modeli na danych z kolejnych lat, rozszerzenie analiz o dodatkowe zmienne (np. meteorologiczne) oraz dalsza ocena zdolności modeli do generalizacji w innych warunkach przestrzennych i czasowych. Praca A3 tworzy może być traktowana zatem jako punkt wyjścia dla kolejnych prezentowanych prac, w których rozwijane są aspekty odporności modeli oraz wpływu dodatkowych danych wejściowych na jakość prognoz.

3.4 Prognozowanie stężeń pyłu zawieszonego z zastosowaniem modelu transformerowego i regresji symbolicznej (artykuł A4)

Celem prac zaprezentowanych w publikacji A4 było sprawdzenie jakości działania hybrydowego modelu prognozowania, integrującego sieć typu *transformer* z regresją symboliczną w zadaniu prognozowania krótkoterminowych stężeń pyłów zawieszonych frakcji PM_{10} oraz $PM_{2,5}$ na podstawie historycznych danych pomiarowych przy dodatkowym wsparciu danych meteorologicznych. Przeprowadzone badania miały także na celu porównanie skuteczności zaproponowanej architektury z klasycznymi metodami prognostycznymi.

W badaniach wykorzystano rzeczywiste dane pochodzące z automatycznych stacji monitoringu jakości powietrza prowadzonych w ramach programu PMŚ z roku 2022. Analiza obejmowała siedem lokalizacji (Zielonkę Bory Tucholskie, Zakopane, Częstochowę, Gliwice/Knurów oraz Szczecin) oraz dwie frakcje pyłów zawieszonych: PM_{10} oraz $PM_{2,5}$. Uwzględnienie różnych lokalizacji pozwoliło ocenić skuteczność modeli w odmiennych warunkach przestrzennych i emisyjnych, a szczególne znaczenie ma również fakt, że badania objęły sezon grzewczy. W badaniach wykorzystano również dane meteorologiczne: temperaturę (°C), prędkość wiatru (km/h) oraz kierunek wiatru (stopnie). Wymienione dane pobrano z platformy Meteostat; dotyczyły tych samych lokalizacji oraz korespondującego okresu pomiarowego.

W publikacji zaproponowano i przetestowano hybrydową architekturę sieci typu *transformer* wraz z modułem regresji symbolicznej. Zaproponowana metoda pozwoliła ocenić, w jakim stopniu zakres informacji historycznej wpływa na jakość krótkoterminowego prognozowania stężeń zanieczyszczeń powietrza. Procedura regresji symbolicznej pozwoliła na znalezienie równania (równanie 3.1) do obliczenia przewidywanego stężenia PM y:

$$y = b_1 \cdot x_1^2 + b_2 \cdot \sin(b_3 \cdot x_2) + b_4 \cdot \cos(b_5 \cdot x_3) \tag{3.1}$$

Współczynniki b_i ($i = 1, \ldots, 5$), które zostały wyznaczone metodą regresji symbolicznej, były równe: $b_1 = -0{,}23$, $b_2 = -0{,}75$, $b_3 = 1{,}12$, $b_4 = 0{,}42$, $b_5 = 0{,}70$. Wartości x_1, x_2 i x_3 odpowiadają następującym danym meteorologicznym: odpowiednio temperaturze, prędkości wiatru i kierunkowi wiatru.

Skuteczność proponowanych modeli porównano z metodami referencyjnymi, obejmującymi następujące podejścia: perceptron wielowarstwowy (MLP), rekurencyjną sieć neuronową (RNN), algorytm ARIMAX, metodę k-NNR oraz metodę LR. Porównanie pozwoliło ocenić, czy proponowana metoda prowadzi do poprawy jakości predykcji. Do oceny jakości prognoz wykorzystano standardowe miary błędu, takie jak MAE, MSE i RMSE. Metoda regresji symbolicznej umożliwia wyznaczenie interpretowalnej formuły opisującej zależności pomiędzy analizowanymi zmiennymi. Poniżej wskazano przykłady, w których zademonstrowano sposób obliczenia prognozy wybranej frakcji PM:

Przykład 1 Wyznaczymy prognozę stężenia PM_{10} na podstawie następujących danych z 4 stycznia 2022 r. na stacji Zielonka Bory Tucholskie (oznaczenie PMŚ: KpZielBoryTu). Znormalizowane wartości temperatury x_1, prędkości wiatru x_2 i kierunku wiatru x_3 przedstawiają się następująco: $x_1 = 0{,}40$, $x_2 = 0{,}39$ i $x_3 = 0{,}65$:

$$y = -0,23 \cdot (0,40)^2 + (-0,75) \cdot \sin(1,12 \cdot 0,39) + 0,42 \cdot \cos(0,70 \cdot 0,65) = 0,19$$

Rzeczywiste (historyczne) stężenie PM_{10} w tym dniu wynosiło 0,20 (znormalizowane). Potwierdza to bardzo wysoką dokładność prognozowania proponowanej metody.

Przykład 2 Znormalizowane wartości temperatury x_1, prędkości wiatru x_2 i kierunku wiatru x_3 z dnia 27 maja (2022 r.) na stacji Szczecinek (oznaczenie PMŚ: ZpSzczec1Maj) są następujące: $x_1 = 0{,}58$, $x_2 = 0{,}41$ i $x_3 = 0{,}72$. Użyjemy tych wartości do oszacowania stężenia $PM_{2,5}$:

$$y = -0,23 \cdot (0,58) + (-0,75) \cdot \sin(1,12 \cdot 0,41) + 0,42 \cdot \cos(0,70 \cdot 0,72) = 0,11$$

Rzeczywiste (historyczne) stężenie $PM_{2,5}$ w tym dniu wynosiło 0,11 (znormalizowane). Potwierdza to wysoką dokładność prognozowania proponowanej hybrydowej metody.

Wyniki uzyskane w pracy A4 wskazują zatem, że proponowana metoda osiąga wysoką dokładność generowanych prognoz. Przykładowo, tabela 3.4 przedstawia wartości miar błędów dla prognoz stężeń PM_{10}.

Tabela 3.4. Metryki błędów proponowanej metody hybrydowej do prognozowania stężeń PM_{10}; wyniki na podstawie tabeli 1 z publikacji A4 [17]

Stacja pomiarowa	Miara	Proponowana	MLP	RNN	ARIMAX	*k*-NNR	LR
KpZielBoryTu	MAE	0,13	0,23	0,21	0,22	0,21	0,22
	MSE	0,03	0,08	0,07	0,08	0,07	0,07
	RMSE	0,16	0,28	0,26	0,27	0,27	0,27
MpZakopaSien	MAE	0,08	0,17	0,17	0,20	0,20	0,20
	MSE	0,01	0,05	0,04	0,06	0,06	0,06
	RMSE	0,11	0,21	0,21	0,25	0,25	0,25
SlCzestoBacz	MAE	0,11	0,22	0,22	0,22	0,21	0,22
	MSE	0,02	0,07	0,08	0,07	0,07	0,07
	RMSE	0,15	0,27	0,28	0,27	0,26	0,27
SlKnurJedNar	MAE	0,10	0,22	0,20	0,22	0,22	0,22
	MSE	0,02	0,08	0,06	0,07	0,07	0,08
	RMSE	0,14	0,28	0,24	0,27	0,27	0,28
ZpSzczec1Maj	MAE	0,12	0,19	0,19	0,23	0,22	0,23
	MSE	0,03	0,06	0,06	0,08	0,07	0,08
	RMSE	0,17	0,24	0,24	0,28	0,27	0,28

Uzyskane wyniki wskazują, że proponowana metoda hybrydowa uzyskała niższe wartości analizowanych miar błędów w rozpatrywanym zbiorze danych. Dla miary MAE wartości błędu mieściły się w zakresie od 0,08 do 0,13, podczas gdy dla metod referencyjnych przyjmowały wartości od 0,17 do 0,23. Podobną zależność zaobserwowano dla miar MSE oraz RMSE. Przykładowo, dla stacji w Zakopanem (MpZakopaSien) wartość RMSE dla proponowanej metody wyniosła 0,11, natomiast dla pozostałych metod osiągała wartości od 0,21 do 0,25. Otrzymane rezultaty wskazują, że zastosowanie regresji symbolicznej pozwoliło na uzyskanie dokładniejszych prognoz stężeń PM_{10} niż w przypadku analizowanych metod referencyjnych. Jednocześnie stosunkowo niewielkie zróżnicowanie wyników pomiędzy poszczególnymi stacjami sugeruje stabilność opracowanego podejścia dla różnych lokalizacji pomiarowych.

Rozszerzenie badań z pracy A4 na lata 2023 i 2024 przedstawiono w rozdz. 4 w tab. 4.3. W tabeli tej pokazano, że po uśrednieniu wyników względem stacji pomiarowych i zanieczyszczeń, proponowana metoda również osiągała niższe wartości błędów niż metody referencyjne. Przykładowo, dla danych z 2022 r. wartość MAE wyniosła 0,11 dla metody proponowanej oraz 0,26 dla metod referencyjnych; dla danych pomiarowych z roku 2023 i 2024, odpowiednio: 0,12 i 0,28 oraz 0,12 i 0,29.

Uzyskane wyniki potwierdzają, że w analizowanym zbiorze danych zastosowanie proponowanej architektury pozwoliło na poprawę jakości krótkoterminowej predykcji względem metod referencyjnych. Poprawa ta była widoczna zarówno w przykładach szczegółowych dla określonych stacji pomiarowych, jak i w wynikach uśrednionych. Oznacza to, że skuteczność zaproponowanych modeli nie ograniczała się wyłącznie do pojedynczych przypadków, lecz była obserwowana również w szerszym zestawieniu wyników.

Wyniki przedstawione w pracy A4 wspierają główną tezę rozprawy pokazując, że przedstawiona hybrydowa architektura może być skutecznym narzędziem krótkoterminowego prognozowania stężeń zanieczyszczeń powietrza w zróżnicowanych warunkach środowiskowych. Podobnie, jak w poprzednich pracach, wkład omawianej pracy polega przede wszystkim na eksperymentalnym potwierdzeniu, że modele typu *data-driven* mogą osiągać lepszą jakość predykcji niż metody referencyjne, a także na pokazaniu, że ocena skuteczności takich modeli powinna uwzględniać lokalizację stacji, czy rodzaj analizowanych zanieczyszczeń.

3.5 Znaczenie danych meteorologicznych w prognozowaniu stężeń pyłu zawieszonego z wykorzystaniem autoenkoderów (artykuł A5)

Celem prac zaprezentowanych w artykule A5 było sprawdzenie, czy nowoczesne modele głębokiego uczenia mogą skutecznie prognozować krótkoterminowe stężenia zanieczyszczeń powietrza na podstawie historycznych danych pomiarowych przy wsparciu danych meteorologicznych. Prowadzone prace miały również na celu porównanie skuteczności zaproponowanych architektur z klasycznymi metodami prognostycznymi.

W badaniach wykorzystano historyczne stężenia pochodzące z automatycznych stacji monitoringu jakości powietrza prowadzonych w ramach programu PMŚ z roku 2022. W analizie obejmującej siedem lokalizacji (Zielonkę Bory Tucholskie, Zakopane, Częstochowę, Gliwice/Knurów oraz Szczecin) uwzględniono dane w zakresie stężeń pyłów zawieszonych frakcji PM_{10} oraz $PM_{2,5}$ i danych meteorologicznych (temperatura (°C), prędkość wiatru (km/h), kierunek wiatru (stopnie) oraz ciśnienie atmosferyczne (hPa)).

W artykule zaproponowano i przetestowano dwie architektury autoenkoderów: konwolucyjno-rekurencyjny (CAE) oraz hierarchiczny (HA). Zastosowanie wymienionych architektur umożliwiło systematyczne porównanie skuteczności modeli w zależności od lokalizacji, rodzaju zanieczyszczenia oraz konfiguracji eksperymentalnej. Uczenie omawianych metod z wykorzystaniem danych meteorologicznych polegało na minimalizacji funkcji straty wyznaczanej pomiędzy rzeczywistymi i prognozowanymi wartościami stężeń zanieczyszczeń. Dane meteorologiczne nie stanowiły zmiennej prognozowanej, lecz pełniły rolę dodatkowych zmiennych wejściowych dostarczających informacji o warunkach atmosferycznych. Dzięki temu modele mogły uwzględniać zależności pomiędzy stężeniami historycznymi a parametrami meteorologicznymi podczas procesu prognozowania. Skuteczność proponowanych modeli porównano z metodą referencyjną – algorytmem RF. Porównanie pozwoliło ocenić, czy proponowana metoda prowadzi do poprawy jakości predykcji. Do oceny jakości prognoz wykorzystano standardowe miary błędu, takie jak MAE i MSE.

Wyniki uzyskane w pracy A5 wskazują, że proponowane metody osiągnęły wysoką dokładność generowanych prognoz, a zastosowanie danych meteorologicznych dodatkowo może prowadzić do podniesienia dokładności prognoz. Przykładowe wyniki zestawiono w tabeli 3.5, która przedstawia wartości miar błędów dla prognoz $PM_{2,5}$.

Wyniki przedstawione w tabeli wskazują, że najlepsze rezultaty uzyskano dla proponowanych metod CAE oraz HA, które w większości analizowanych przypadków osiągały najniższe wartości błędów MAE i MSE. Szczególnie korzystne wyniki zaobserwowano dla stacji Zakopane, gdzie wartość MAE wyniosła odpowiednio 0,11 i 0,12, natomiast MSE osiągnęło poziom 0,02 i 0,03. Metody nie wykorzystujące danych meteorologicznych (oznaczone gwiazdką) uzyskiwały zazwyczaj wyższe wartości błędów. W większości przypadków również model RF (w obu konfiguracjach) osiągał wyższe wartości błędów względem proponowanych metod. Otrzymane rezultaty wskazują ponadto, że przewaga metod CAE i HA względem rozwiązań algorytmu RF utrzymywała się dla wszystkich analizowanych stacji pomiarowych, co świadczy o stabilności uzyskiwanych wyników w analizowanym zbiorze danych.

Rozszerzenie badań z pracy A5 na lata 2023 i 2024 przedstawiono w rozdz. 4 w tab. 4.4. Po uśrednieniu wyników względem stacji pomiarowych i zanieczyszczeń, omawiane metody również

Tabela 3.5. Porównanie wartości MAE oraz MSE dla analizowanych stacji pomiarowych dla prognoz stężeń $PM_{2,5}$; wyniki na podstawie tabeli 1 z publikacji A5 [14]

Stacja pomiarowa	Miara	CAE	HA	CAE*	HA*	RF	RF*
Zielonka Bory Tucholskie	MAE	0,150	0,160	0,230	0,230	0,190	0,210
	MSE	0,040	0,040	0,120	0,110	0,070	0,090
Zakopane	MAE	0,110	0,120	0,190	0,190	0,210	0,230
	MSE	0,020	0,030	0,100	0,100	0,060	0,080
Częstochowa	MAE	0,210	0,190	0,290	0,260	0,220	0,240
	MSE	0,060	0,070	0,140	0,140	0,100	0,120
Gliwice	MAE	0,250	0,210	0,330	0,280	0,240	0,260
	MSE	0,090	0,130	0,170	0,200	0,190	0,210
Szczecin	MAE	0,200	0,220	0,280	0,290	0,250	0,270
	MSE	0,070	0,100	0,150	0,170	0,130	0,150

osiągnęły niższe wartości błędów niż metoda referencyjna. Na przykład, dla danych z 2022 r. wartość MAE wyniosła 0,11 i 0,09 dla metod proponowanych oraz 0,26 i 0,21 dla metody referencyjnej (odpowiednio, bez uwzględnienia oraz z uwzględnieniem danych meteorologicznych). Dla danych pomiarowych z roku 2023 uzyskano wyniki: 0,12 i 0,10 (metody proponowane) oraz 0,28 i 0,24 (metoda bazowa); dla roku 2024: 0,12 i 0,10 (metody proponowane) oraz 0,28 i 0.24 (metoda bazowa).

Uzyskane wyniki potwierdzają, że w analizowanym zbiorze danych zastosowanie proponowanych architektur pozwoliło na poprawę jakości krótkoterminowej predykcji względem metody referencyjnej. Poprawa ta była widoczna zarówno w przykładach szczegółowych dla określonych stacji pomiarowych, jak i w wynikach uśrednionych. Oznacza to, że skuteczność zaproponowanych modeli nie ograniczała się wyłącznie do pojedynczych przypadków, lecz była obserwowana również w szerszym zestawieniu wyników.

Wyniki przeprowadzonych badań w pracy A5 wspierają tezę rozprawy wskazując, że przedstawione architektury oparte na metodach uczenia głębokiego mogą być skutecznym narzędziem krótkoterminowego prognozowania stężeń pyłów zawieszonych w zróżnicowanych warunkach środowiskowych. Podobnie, jak w poprzednich pracach, wkład omawianego artykułu polega przede wszystkim na eksperymentalnym potwierdzeniu, że modele typu *data-driven* mogą osiągać lepszą jakość predykcji niż metody referencyjne, a także na pokazaniu, że ocena skuteczności takich modeli powinna uwzględniać lokalizację stacji, czy rodzaj analizowanych zanieczyszczeń.

3.6 Dokładność i wyjaśnialność prognozowania stężeń pyłu zawieszonego z wykorzystaniem sieci neuronowych i danych meteorologicznych (publikacja konferencyjna A6)

Publikacja konferencyjna A6 jest kolejną pracą badawczą, w której zaprezentowano wyniki przeprowadzonych analiz i obliczeń mających na celu ocenę nowoczesnych modeli głębokiego uczenia do prognozowania krótkoterminowych stężeń zanieczyszczeń powietrza na podstawie historycznych danych pomiarowych oraz danych meteorologicznych. Badania miały także na celu porównanie skuteczności zaproponowanych architektur z klasycznymi metodami prognostycznymi.

W badaniach wykorzystano historyczne stężenia pochodzące z automatycznych stacji monitoringu jakości powietrza prowadzonych w ramach programu PMŚ z roku 2022. Analiza obejmowała siedem lokalizacji (Zielonkę Bory Tucholskie, Zakopane, Częstochowę, Gliwice/Knurów oraz Szczecin) oraz uwzględniała dane historyczne dotyczące frakcji pyłów zawieszonych: PM_{10} oraz $PM_{2,5}$ oraz dane meteorologiczne: temperaturę (°C), prędkość wiatru (km/h) i kierunek wiatru (stopnie).

W artykule przeanalizowano trzy architektury: model *temporal fusion transformer*, sieć Kołmogorowa–Arnolda oraz sieć konwolucyjną temporalną (TCN). Zastosowanie wymienionych modeli umożliwiło systematyczne porównanie ich skuteczności w zależności od lokalizacji, rodzaju zanieczyszczenia oraz konfiguracji eksperymentalnej. W przypadku eksperymentów uwzględniających dane meteorologiczne proces uczenia opierał się na minimalizacji błędu pomiędzy wartościami rzeczywistymi i prognozowanymi stężeń zanieczyszczeń. Parametry meteorologiczne pełniły rolę dodatkowego kontekstu opisującego warunki atmosferyczne. Skuteczność proponowanych modeli porównano z metodami referencyjnymi – rekurencyjną siecią neuronową (RNN) oraz siecią długiej pamięci krótkotrwałej (LSTM). Porównanie pozwoliło ocenić, czy dyskutowane metody prowadzą do poprawy jakości predykcji. Do oceny jakości prognoz wykorzystano standardowe miary błędu, takie jak MAE, RMSE oraz R^2.

Wyniki zaprezentowane w publikacji A6 wskazują, że omawiane modele głębokiego uczenia osiągały niższe wartości miar błędu niż metody referencyjne w rozważanym zbiorze danych. Przykładowe wyniki zestawiono w tabeli 3.6, która przedstawia wartości miar błędów dla prognoz stężeń PM_{10}.

Analiza wyników wskazuje, że uwzględnienie danych meteorologicznych prowadzi do poprawy jakości prognoz stężeń PM_{10} dla wszystkich analizowanych stacji pomiarowych i większości

Tabela 3.6. Porównanie wartości miar błędów dla analizowanych stacji pomiarowych dla prognoz stężeń PM_{10} bez oraz z uwzględnieniem danych meteorologicznych (oznaczone gwiazdką); wyniki na podstawie tabeli 2 z pracy A6 [20]

	Miara	TFT	TFT*	KAN	KAN*	TCN	TCN*	RNN	RNN*	LSTM	LSTM*
	MAE	0,11	0,10	0,14	0,09	0,12	0,10	0,22	0,21	0,18	0,16
KpZielBoryTu	RMSE	0,25	0,22	0,26	0,22	0,26	0,21	0,29	0,28	0,26	0,25
	R^2	0,88	0,91	0,88	0,91	0,87	0,90	0,84	0,85	0,85	0,86
	MAE	0,12	0,10	0,13	0,10	0,14	0,09	0,22	0,20	0,23	0,21
MpZakopaSien	RMSE	0,26	0,21	0,25	0,21	0,25	0,22	0,26	0,25	0,25	0,24
	R^2	0,87	0,91	0,85	0,91	0,87	0,92	0,83	0,85	0,84	0,86
	MAE	0,11	0,09	0,10	0,10	0,10	0,09	0,18	0,16	0,16	0,14
SlCzestoBacz	RMSE	0,26	0,21	0,24	0,21	0,26	0,23	0,27	0,25	0,26	0,24
	R^2	0,89	0,89	0,84	0,89	0,85	0,91	0,82	0,83	0,84	0,85
	MAE	0,12	0,09	0,15	0,09	0,13	0,08	0,18	0,16	0,15	0,14
SlKnurJedNar	RMSE	0,26	0,22	0,29	0,22	0,24	0,21	0,29	0,25	0,26	0,22
	R^2	0,88	0,90	0,87	0,91	0,87	0,92	0,82	0,84	0,85	0,87
	MAE	0,13	0,11	0,11	0,09	0,14	0,11	0,17	0,15	0,16	0,13
ZpSzczec1Maj	RMSE	0,25	0,21	0,24	0,22	0,26	0,23	0,28	0,26	0,24	0,23
	R^2	0,87	0,89	0,85	0,91	0,84	0,90	0,81	0,85	0,84	0,86

badanych architektur sieci neuronowych. Poprawa ta przejawia się obniżeniem wartości błędów MAE i RMSE oraz wzrostem współczynnika determinacji R^2. Największe korzyści z zastosowania dodatkowych zmiennych meteorologicznych zaobserwowano dla modeli KAN i TCN, które osiągały najniższe wartości błędów, przykładowo dla stacji Knurów (oznaczenie PMŚ: SlKnurJedNar) model TCN* uzyskał MAE równe 0,08 oraz R^2 na poziomie 0,92. Modele TFT również wykazały wyraźną poprawę skuteczności, osiągając wartości R^2 przekraczające 0,90 dla większości lokalizacji. Z kolei modele bazowe RNN i LSTM charakteryzowały się najwyższymi błędami prognozowania zarówno bez, jak i z uwzględnieniem danych meteorologicznych. Najlepsze rezultaty uzyskano dla modeli TCN* i KAN*, które najczęściej osiągały najniższe wartości MAE i RMSE oraz najwyższe wartości współczynnika R^2.

Badania rozszerzone na lata 2023 i 2024 przedstawiono w rozdz. 4 w tab. 4.5. Po uśrednieniu wyników względem stacji pomiarowych i zanieczyszczeń, omawiane metody również osiągnęły niższe wartości błędów niż metody referencyjne. Na przykład, dla danych z 2022 r. wartość MAE wyniosła 0,11 i 0,09 dla metod proponowanych oraz 0,26 i 0,21 dla metod referencyjnych (odpowiednio, bez uwzględnienia oraz z uwzględnieniem danych meteorologicznych). Dla danych pomiarowych z roku 2023 uzyskano wyniki: 0,12 i 0,10 (metody proponowane) oraz 0,28 i 0,24 (metody bazowe); dla roku 2024: 0,12 i 0,10 (metody proponowane) oraz 0,28 i 0.24

(metody bazowe).

Przeprowadzona analiza wykazała, że zastosowanie proponowanych architektur sieci neuronowych umożliwiło uzyskanie wyższej jakości krótkoterminowych prognoz frakcji pyłu zawieszonego w porównaniu z metodami referencyjnymi. Korzystniejsze wyniki odnotowano zarówno dla wybranych stacji pomiarowych analizowanych indywidualnie, jak i w przypadku wartości uśrednionych dla całego analizowanego zbioru danych. Wskazuje to, że poprawa skuteczności modeli miała charakter systematyczny i nie była ograniczona do pojedynczych lokalizacji. Ponadto zaobserwowano, że rozszerzenie danych wejściowych o parametry meteorologiczne przyczynia się do dalszego zwiększenia dokładności generowanych prognoz.

Zaprezentowane w pracy A6 wyniki wspierają tezę rozprawy wskazując, że przedstawione architektury oparte na metodach uczenia głębokiego mogą być skutecznym narzędziem krótkoterminowego prognozowania stężeń pyłów zawieszonych w zróżnicowanych warunkach środowiskowych. Podobnie, jak w przypadku wcześniejszych prac, jako ograniczenie pracy A6 można wskazać, że ocena proponowanego rozwiązania została przeprowadzona dla określonych zbiorów danych, lokalizacji pomiarowych oraz przyjętego schematu eksperymentalnego. Powoduje to, że uzyskane rezultaty należy rozważać w kontekście analizowanych danych oraz zastosowanego schematu walidacji.

3.7 Ocena skuteczności modeli głębokiego uczenia w krótkoterminowym prognozowaniu stężeń pyłu zawieszonego z uwzględnieniem danych meteorologicznych (artykuł A7)

Ostatnim artykułem, w którym zaprezentowano wyniki badań będących rozszerzeniem analiz i rezultatów uzyskanych w publikacji A6, jest artykuł A7. W publikacji tej rozszerzono analizę o dane pomiarowe z lat 2023 i 2024. Celem prowadzonych prac jest zatem ocena nowoczesnych modeli głębokiego uczenia do prognozowania krótkoterminowych stężeń zanieczyszczeń powietrza na podstawie historycznych danych pomiarowych oraz danych meteorologicznych. Ponadto, wyniki omawianych architektur zostały porównane z siecią LSTM.

W tabeli 3.7 przedstawiono wartości miar błędów dla prognoz stężeń PM_{10}.

Analiza uzyskanych wyników wskazuje, że modele trenowane z zarówno stężeniami historycznymi, jak i zmiennymi meteorologicznymi osiągały mniejsze błędy prognoz w analizowanym zbiorze danych na wszystkich stacjach pomiarowych. Przykładowo, model TFT* dla stacji Zako-

Tabela 3.7. Wyniki prognoz stężeń PM_{10}. Gwiazdka oznacza, że do treningu modeli wykorzystano zarówno stężenia historyczne, jak i zmienne meteorologiczne; wyniki na podstawie tabeli 2 z publikacji A7 [19]

Stacja pomiarowa	Miara błędu	KAN	KAN*	TFT	TFT*	LSTM	LSTM*
Zielonka Bory Tucholskie	MAE	0,13	0,11	0,14	0,11	0,17	0,16
	MSE	0,04	0,03	0,06	0,04	0,07	0,06
	R^2	0,87	0,90	0,87	0,89	0,85	0,86
Zakopane	MAE	0,12	0,11	0,15	0,12	0,20	0,15
	MSE	0,05	0,04	0,04	0,03	0,08	0,05
	R^2	0,86	0,89	0,87	0,91	0,84	0,85
Częstochowa	MAE	0,11	0,11	0,15	0,11	0,18	0,17
	MSE	0,05	0,03	0,05	0,03	0,07	0,05
	R^2	0,87	0,89	0,87	0,90	0,85	0,86
Gliwice	MAE	0,14	0,12	0,15	0,11	0,18	0,17
	MSE	0,05	0,04	0,04	0,03	0,06	0,06
	R^2	0,86	0,89	0,87	0,89	0,84	0,86
Szczecin	MAE	0,14	0,11	0,15	0,12	0,19	0,15
	MSE	0,04	0,03	0,05	0,03	0,08	0,06
	R^2	0,87	0,90	0,88	0,90	0,84	0,86

pane uzyskał najwyższą wartość współczynnika R^2 równą 0,91. Model LSTM osiągnął najwyższe wartości błędów spośród rozważanych podejść. Wyniki wydają się być spójne geograficznie, tzn. żadna ze stacji nie różni się istotnie od pozostałych, co sugeruje, że zaproponowane modele dobrze generalizują się na różne lokalizacje w Polsce.

Przeprowadzona analiza wykazała, że zastosowanie proponowanych architektur sieci neuronowych umożliwiło uzyskanie wyższej jakości krótkoterminowych prognoz frakcji pyłu zawieszonego w porównaniu z metodami referencyjnymi. Korzystniejsze wyniki odnotowano zarówno dla wybranych stacji pomiarowych analizowanych indywidualnie, jak i w przypadku wartości uśrednionych dla całego analizowanego zbioru danych. Wskazuje to, że poprawa skuteczności modeli miała charakter systematyczny i nie była ograniczona do pojedynczych lokalizacji. Ponadto zaobserwowano, że rozszerzenie danych wejściowych o parametry meteorologiczne przyczynia się do dalszego zwiększenia dokładności generowanych prognoz.

Podsumowując, wyniki przedstawione w pracy A7 stanowią potwierdzenie tezy rozprawy, wskazując, że architektury bazujące na metodach głębokiego uczenia mogą efektywnie wspierać krótkoterminowe prognozowanie stężeń pyłów zawieszonych w różnych warunkach środowiskowych. Jednocześnie, podobnie jak w przypadku wcześniej omówionych publikacji, należy uwzględnić ograniczenia badania wynikające z faktu, że weryfikacja skuteczności zaproponowanego podejścia została przeprowadzona na wybranych zbiorach danych oraz dla konkretnych lokalizacji pomiarowych. W konsekwencji uzyskane wyniki powinny być interpretowane w odniesieniu do analizowanych danych, przyjętych warunków eksperymentalnych oraz zastosowanej procedury walidacyjnej.

Podsumowując, publikacja A7 stanowi istotny element części eksperymentalnej rozprawy, w której przeprowadzono autorskie badania z wykorzystaniem rzeczywistych danych pomiarowych. W pracy dokonano oceny zaproponowanych metod poprzez ich porównanie z klasycznym modelem referencyjnym, uzyskując wyniki bezpośrednio potwierdzające założenia i tezę rozprawy. Wartość naukowa pracy A7 wynika zatem zarówno z osiągnięcia lepszych wyników prognostycznych wyrażonych niższymi wartościami miar błędów, jak również z wykazania konieczności analizowania problemu prognozowania jakości powietrza z uwzględnieniem wielu lokalizacji, różnych rodzajów zanieczyszczeń oraz zmienności sezonowej.

3.8 Podsumowanie

Przedstawione w niniejszym rozdziale wybrane wyniki badań i analiz, zaprezentowane w pracach naukowych A1–A7, tworzą spójną podstawę merytoryczną rozprawy doktorskiej i są bezpośrednio powiązane z jej celem oraz postawioną tezą badawczą. Przeprowadzone eksperymenty wykazały, że metody oparte na uczeniu głębokim mogą skutecznie wspierać krótkoterminowe prognozowanie stężeń zanieczyszczeń powietrza, w szczególności pyłu zawieszonego (frakcja PM_{10} i $PM_{2,5}$). Zaproponowane architektury, obejmujące zarówno klasyczne sieci rekurencyjne, jak i nowoczesne modele (m.in. TCN, FCN, GRU, xLSTM, KAN, VAE, TFT), osiągały wyniki porównywalne lub lepsze od modeli referencyjnych, co znajdowało odzwierciedlenie w niższych wartościach miar błędów. Uzyskane rezultaty potwierdzają konieczność prowadzenia analiz z uwzględnieniem wielu lokalizacji pomiarowych oraz możliwie jak najszerszej informacji, w tym w zakresie zmienności analizowanych stężeń zanieczyszczeń oraz towarzyszącym zanieczyszczeniu warunkom meteorologicznym. Jednocześnie należy podkreślić, że przedstawione wyniki zostały uzyskane dla określonych zbiorów danych i wybranych stacji pomiarowych, przy zastosowaniu konkretnych procedur podziału danych oraz walidacji modeli. Oznacza to, że choć otrzymane rezultaty potwierdzają wysoką skuteczność proponowanych rozwiązań, ich interpretacja powinna uwzględniać charakter analizowanych danych i warunki eksperymentalne. W konsekwencji, dalsze badania obejmujące większą liczbę lokalizacji, dłuższe szeregi czasowe oraz dane pochodzące z innych regionów mogłyby dostarczyć dodatkowych przesłanek dotyczących możliwości generalizacji zaproponowanych metod i ich zastosowania w szerszym zakresie praktycznych systemów monitorowania oraz prognozowania jakości powietrza.

Rozdział 4

Uzupełniająca analiza wyników badań przedstawionych w pracach A2–A6

Wyniki przedstawione w poprzednim rozdziale obejmowały rezultaty badań zaprezentowanych w pracach A2–A7. W przypadku prac A2–A6 podstawę analiz stanowiły dane PMŚ z roku 2022, natomiast praca A7 została oparta na nowszym zakresie danych, obejmującym lata 2023–2024. W związku z tym, w niniejszym rozdziale przedstawiono uzupełniające eksperymenty odnoszące się do metod zaproponowanych w pracach A2–A6, przeprowadzone z wykorzystaniem zweryfikowanych danych historycznych PMŚ z lat 2023 i 2024. Ich celem była ocena stabilności oraz zdolności uogólniania opracowanych modeli w odniesieniu do nowszych obserwacji, które nie były dostępne na etapie prowadzenia pierwotnych badań dla prac A2–A6.

Eksperymenty dla danych z lat 2023 i 2024, tj. danych zweryfikowanych i opublikowanych przez GIOŚ, przeprowadzono w sposób analogiczny do analiz przedstawionych w pracach A2–A6. Dane meteorologiczne ponownie pobrano z platformy Meteostat. Ze względu na rozległość przeprowadzonych eksperymentów oraz potrzebę zachowania przejrzystości prezentacji, wyniki uśredniono z uwzględnieniem wszystkich rozpatrywanych stacji pomiarowych, analizowanych zanieczyszczeń oraz okresów badawczych. Eksperymenty powtórzono z zastosowaniem konfiguracji sieci zgodnych z podejściami omówionymi w rozdz. 3 dla prac A2–A6.

4.1 Prognozowanie stężeń zanieczyszczeń powietrza (artykuły A2 i A3)

W ramach zaprezentowanych prac A2 oraz A3 zaproponowano sieci TCN, KAN, FCN, GRU oraz autoenkoder wariacyjny (VAE), które zostały porównane z metodami: wygładzania wykładniczego, wygładzania wykładniczego Holta-Wintersa, algorytmami ARIMA oraz RF. Wyniki przedstawiono w tabelach 4.1 oraz 4.2.

Tabela 4.1. Dokładność prognozowania na podstawie stężeń historycznych – wyniki uśrednione (ze względu na stacje pomiarowe oraz zanieczyszczenia), artykuł A2

Rok	Typ danych	Miara	Proponowane	Metody bazowe
2022	godzinowe	MAE	0,11	0,19
		MAPE	14,25	37,25
		MSE	0,02	0,08
		RMSE	0,14	0,28
	dobowe	MAE	0,11	0,22
		MAPE	16,61	41,26
		MSE	0,03	0,10
		RMSE	0,17	0,32
2023	godzinowe	MAE	0,12	0,21
		MAPE	15,11	39,61
		MSE	0,02	0,09
		RMSE	0,14	0,30
	dobowe	MAE	0,13	0,24
		MAPE	17,15	43,36
		MSE	0,04	0,11
		RMSE	0,20	0,33
2024	godzinowe	MAE	0,13	0,21
		MAPE	15,79	41,42
		MSE	0,03	0,11
		RMSE	0,17	0,33
	dobowe	MAE	0,12	0,24
		MAPE	18,19	46,81
		MSE	0,04	0,12
		RMSE	0,20	0,35

Wyniki wskazują, że skuteczność proponowanej metody utrzymuje się na zbliżonym poziomie również dla nowszych danych obejmujących lata 2023 i 2024. Dla danych godzinowych wartości błędu MAE dla proponowanej metody wyniosły odpowiednio 0,11, 0,12 oraz 0,13 dla analizowanych lat, natomiast wartości RMSE przyjmowały wartości 0,14 i 0,17. Analogicznie,

Tabela 4.2. Dokładność prognozowania na podstawie stężeń historycznych – wyniki uśrednione (ze względu na stacje pomiarowe oraz zanieczyszczenia), artykuł A3

Rok	Typ danych	Miara	Proponowane	Metody bazowe
2022	godzinowe	MAE	0,11	0,19
		MSE	0,02	0,08
		RMSE	0,14	0,28
	dobowe	MAE	0,11	0,22
		MSE	0,03	0,10
		RMSE	0,17	0,32
2023	godzinowe	MAE	0,12	0,21
		MSE	0,03	0,11
		RMSE	0,17	0,33
	dobowe	MAE	0,12	0,24
		MSE	0,03	0,11
		RMSE	0,17	0,33
2024	godzinowe	MAE	0,12	0,21
		MSE	0,03	0,12
		RMSE	0,18	0,34
	dobowe	MAE	0,11	0,24
		MSE	0,03	0,13
		RMSE	0,17	0,36

dla danych dobowych uzyskano wartości MAE w zakresie od 0,11 do 0,13 oraz RMSE od 0,17 do 0,20. Otrzymane rezultaty wskazują, że różnice pomiędzy poszczególnymi latami są niewielkie i nie wskazują na istotne pogorszenie jakości prognoz wraz z upływem czasu.

Jednocześnie w przeprowadzonych eksperymentach proponowane rozwiązanie uzyskało niższe wartości błędów niż metody referencyjne opisane w literaturze, dla wszystkich analizowanych lat i rozważanych miar jakości. Przykładowo, w przypadku danych godzinowych w roku 2024 wartość MAE dla metod literaturowych wyniosła 0,21, podczas gdy dla proponowanego rozwiązania osiągnęła wartość 0,13. Podobną tendencję zaobserwowano dla pozostałych analizowanych miar, tj. MAPE, MSE oraz RMSE. Warto również zauważyć, że względna różnica pomiędzy proponowaną metodą a rozwiązaniami referencyjnymi pozostawała na zbliżonym poziomie dla poszczególnych lat, co wskazuje na stabilne zachowanie opracowanego podejścia w analizowanym zbiorze danych.

Wyniki przedstawione w tab. 4.2 są zgodne z obserwacjami uzyskanymi dla danych z roku 2022 również w odniesieniu do lat 2023 i 2024. Wartości wszystkich analizowanych miar błędu dla proponowanych metod utrzymują się na zbliżonym poziomie. Dla danych godzinowych wartość MAE wzrosła nieznacznie z 0,11 w roku 2022 do 0,12 w latach 2023 i 2024, natomiast RMSE wzrosło z 0,14 do 0,18. Analogiczne zależności zaobserwowano dla danych dobowych, dla których uzyskano zbliżone wartości miar błędu w latach 2023 i 2024.

Jednocześnie w przeprowadzonych eksperymentach proponowane rozwiązanie uzyskało niższe wartości błędów niż metody referencyjne dla wszystkich analizowanych przypadków. Przykładowo, dla danych godzinowych w roku 2024 wartość RMSE wyniosła 0,18 dla proponowanych metod oraz 0,34 dla metod literaturowych, co wskazuje na około dwukrotnie mniejszy błąd prognozowania. Otrzymane rezultaty wskazują, że skuteczność opracowanych modeli utrzymuje się również dla danych pochodzących z okresu późniejszego niż wykorzystywany w badaniach pierwotnych, co sugeruje ich zdolność do generalizacji na nowe dane.

Uzyskane wyniki potwierdzają skuteczność w analizowanym zbiorze danych z lat 2023 i 2024. Tym samym przeprowadzone eksperymenty wzmacniają przesłanki dotyczące stabilności modeli oraz ich zdolności do skutecznego prognozowania wartości stężeń zanieczyszczeń powietrza w dłuższym horyzoncie czasowym.

4.2 Prognozowanie stężeń pyłu zawieszonego (artykuły A4–A6)

Artykuł A4 W pracy A4 zaproponowano model sieci neuronowej typu *transformer* połączony z modułem regresji symbolicznej, który został porównany z metodami takimi, jak: sieć MLP i RNN, algorytm ARIMAX, *k*-NNR oraz LR. Wyniki przedstawiono w tabeli 4.3.

Tabela 4.3. Dokładność prognozowania na podstawie stężeń historycznych – wyniki dobowe uśrednione (ze względu na stacje pomiarowe oraz zanieczyszczenia), artykuł A4

Rok	Miara	Proponowana	Metody bazowe
2022	MAE	0,11	0,26
	MSE	0,03	0,52
	RMSE	0,17	0,72
2023	MAE	0,12	0,28
	MSE	0,03	0,53
	RMSE	0,17	0,73
2024	MAE	0,12	0,29
	MSE	0,04	0,53
	RMSE	0,20	0,73

Wyniki wskazują, że proponowana metoda uzyskuje w analizowanym zbiorze danych niższe wartości analizowanych miar błędów niż metody bazowe. Przykładowo, dla roku 2022 wartość MAE wyniosła 0,11 dla proponowanego rozwiązania oraz 0,26 dla metod bazowych, natomiast RMSE osiągnęło odpowiednio 0,17 i 0,72. Podobne zależności zaobserwowano również dla danych z lat 2023 i 2024. Jednocześnie wartości błędów dla proponowanej metody pozostają na zbliżonym poziomie we wszystkich analizowanych latach, przy niewielkim wzroście MAE z 0,11 do 0,12 oraz RMSE z 0,17 do 0,20. Otrzymane rezultaty wskazują, że skuteczność opracowanego podejścia utrzymuje się również dla danych pochodzących z kolejnych lat pomiarowych, a uzyskana poprawa względem metod bazowych pozostaje wyraźna niezależnie od analizowanego okresu.

Artykuł A5 Artykuł A5 przedstawiał autoenkodery hierarchiczne (HA) oraz konwolucyjno-rekurencyjny (CAE), które zostały odniesione do algorytmu RF. Wyniki badań zostały zaprezentowane w tabeli 4.4.

Tabela 4.4. Dokładność prognozowania na podstawie stężeń historycznych (a) oraz z uwzględnieniem danych meteorologicznych (b) – wyniki dobowe uśrednione (ze względu na stacje pomiarowe oraz zanieczyszczenia), artykuł A5

		Proponowane		Metoda bazowa	
Rok	Miara	(a)	(b)	(a)	(b)
2022	MAE	0,11	0,09	0,26	0,21
	MSE	0,03	0,03	0,56	0,52
2023	MAE	0,12	0,10	0,28	0,24
	MSE	0,04	0,03	0,61	0,58
2024	MAE	0,12	0,10	0,28	0,24
	MSE	0,04	0,03	0,59	0,53

Analiza wskazuje, że skuteczność proponowanych metod utrzymuje się na zbliżonym poziomie dla danych z lat 2023 i 2024. Uzyskane wartości miar MAE i MSE są bardzo zbliżone do wyników osiągniętych dla roku 2022, co wskazuje na stabilność opracowanych podejść.

Podobnie, jak w przypadku wyników uzyskanych dla roku 2022, uwzględnienie danych meteorologicznych prowadzi do zmniejszenia błędu MAE z około 0,12 do 0,10 dla lat 2023 i 2024. Jednocześnie, uzyskano niższe wartości błędów proponowanych metod nad rozwiązaniami literaturowymi pozostaje wyraźna we wszystkich analizowanych latach. Otrzymane rezultaty wskazują, że zaproponowane modele wykazują stabilność działania również dla nowszych danych pomiarowych.

Artykuł A6 W artykule A6 zaprezentowano warianty sieci KAN, TCN oraz TFT, których rezultaty zostały porównane z sieciami RNN i LSTM. Rezultaty eksperymentów prognostycznych przedstawiono w tabeli 4.5.

Ponownie, uzyskane wyniki sugerują, że skuteczność proponowanych metod utrzymuje się na wysokim poziomie również dla danych z lat 2023 i 2024. Wartości błędów MAE i RMSE pozostają bardzo zbliżone do wyników uzyskanych dla roku 2022, co wskazuje na stabilność modeli. Przykładowo, współczynnik determinacji (R^2) osiąga wartości z przedziału 0,83–0,85 dla proponowanych rozwiązań, podczas gdy metody referencyjne uzyskują wartości około 0,59–0,66. Wskazuje to, że opracowane modele znacznie lepiej odwzorowują rzeczywistą zmienność prognozowanych stężeń zanieczyszczeń.

Podobnie jak w poprzednich eksperymentach, uwzględnienie danych meteorologicznych na ogół prowadzi do poprawy uzyskiwanych wyników. Zarówno wartości miar błędów, jak i współ-

Tabela 4.5. Dokładność prognozowania na podstawie stężeń historycznych (a) oraz z uwzględnieniem danych meteorologicznych (b) – wyniki dobowe uśrednione (ze względu na stacje pomiarowe oraz zanieczyszczenia), artykuł A6

		Proponowane		Metody bazowe	
Rok	Miara	(a)	(b)	(a)	(b)
2022	MAE	0,11	0,09	0,26	0,21
	RMSE	0,17	0,17	0,72	0,75
	R^2	0,83	0,85	0,59	0,61
2023	MAE	0,12	0,10	0,28	0,24
	RMSE	0,17	0,20	0,73	0,78
	R^2	0,83	0,84	0,62	0,66
2024	MAE	0,12	0,10	0,28	0,24
	RMSE	0,20	0,17	0,77	0,73
	R^2	0,84	0,85	0,61	0,63

czynnika determinacji pozostają na zbliżonym poziomie we wszystkich analizowanych latach. Uzyskane rezultaty wskazują, że wnioski sformułowane na podstawie danych z roku 2022 pozostają aktualne również dla danych z lat 2023 i 2024. Jednocześnie wyniki sugerują zdolność proponowanych metod do uzyskiwania stabilnych prognoz dla nowych danych pomiarowych.

4.3 Podsumowanie

Podsumowując, dodatkowe eksperymenty przeprowadzone dla danych PMŚ oraz meteorologicznych z lat 2023 i 2024 są zgodne z wnioskami sformułowanymi na podstawie wyników uzyskanych dla roku 2022. W analizowanych przypadkach proponowane metody utrzymały korzystne właściwości prognostyczne oraz uzyskały niższe wartości błędów w porównaniu z rozwiązaniami referencyjnymi opisanymi w literaturze. Zaobserwowane różnice wartości poszczególnych miar błędów pomiędzy kolejnymi latami były niewielkie, co wskazuje na stabilne zachowanie opracowanych modeli oraz ich możliwość generalizacji na dane pochodzące z kolejnych okresów pomiarowych.

Rozdział 5

Podsumowanie

W niniejszej rozprawie poruszono tematykę związaną z metodami oceny jakości powietrza, ze szczególnym uwzględnieniem zagadnień związanych z prognozowaniem stężeń wybranych zanieczyszczeń atmosferycznych.

W ramach przeprowadzonych prac zaproponowano i poddano szczegółowej analizie zestaw metod opartych na technikach głębokiego uczenia, przeznaczonych do prognozowania krótkoterminowych stężeń zanieczyszczeń powietrza. Rozważono architektury umożliwiające efektywne modelowanie zależności czasowych w szeregach danych, w tym bramkowe sieci rekurencyjne typu GRU, sieci w pełni konwolucyjne (FCN), sieci temporalne, a także nowoczesne podejścia bazujące na sieciach Kołmogorowa–Arnolda (KAN), sieciach typu transformer oraz autoenkoderach przystosowanych do zadań predykcyjnych. Modele trenowano zarówno w wariantach wykorzystujących wyłącznie historyczne stężenia zanieczyszczeń, jak i w rozszerzonych konfiguracjach uwzględniających dane meteorologiczne. Przeprowadzone porównania wykazały, że rozszerzenie przestrzeni cech o zmienne meteorologiczne może prowadzić do poprawy jakości predykcji w analizowanym zbiorze danych. Uzyskane wyniki wskazują jednocześnie, że końcowa skuteczność prognozowania zależy zarówno od zastosowanej architektury modeli, jak i od zakresu dostępnych danych wejściowych.

Zrealizowane badania potwierdziły skuteczność zaproponowanych metod w analizowanym zbiorze danych obejmującym wybrane lokalizacje pomiarowe w Polsce. Modele osiągnęły niskie wartości miar błędów predykcji, takich jak MAE, MSE oraz RMSE. Przeprowadzona analiza statystyczna wykazała istotną przewagę opracowanych podejść nad klasycznymi metodami statystycznymi oraz wybranymi architekturami referencyjnymi w badanej próbie, co potwierdziło zasadność zastosowania metod głębokiego uczenia w zadaniach prognozowania jakości powietrza w rozpatrywanych warunkach. Istotnym elementem badań było wykorzystanie danych pochodzących ze stacji pomiarowych GIOŚ, zlokalizowanych w zróżnicowanych warunkach środowiskowych, obejmujących zarówno obszary metropolitalne, miasta średniej wielkości, jak i miejscowości uzdrowiskowe oraz obszary o niższym stopniu urbanizacji. Dobór danych umożliwił ocenę odporności modeli w analizowanym zbiorze danych oraz ich zdolności do generalizacji pomiędzy rozpatrywanymi lokalizacjami.

Uzyskane wyniki badań potwierdziły zasadność postawionej tezy, wskazując, że metody głębokiego uczenia zapewniają wyższą dokładność prognozowania stężeń pyłów zawieszonych oraz innych zanieczyszczeń powietrza niż klasyczne podejścia statystyczne i regresyjne w zróżnicowanych warunkach geograficznych i środowiskowych dla analizowanego zbioru danych. W konsekwencji, zrealizowano cel rozprawy, który polegał na opracowaniu i ocenie nowoczesnych metod prognozowania stężeń zanieczyszczeń powietrza w Polsce z uwzględnieniem zróżnicowanych warunków geograficznych i środowiskowych. Zrealizowane zostały również wszystkie założone zadania badawcze. Za najistotniejsze, oryginalne osiągnięcia niniejszej rozprawy można uznać:

(1) Opracowanie i ocenę metod prognozowania stężeń zanieczyszczeń powietrza opartych na technikach głębokiego uczenia, testowanych w analizowanym zbiorze danych zarówno w wariantach wykorzystujących historyczne stężenia zanieczyszczeń, jak i w rozszerzonych konfiguracjach uwzględniających dodatkowo dane meteorologiczne;

(2) Przeprowadzenie eksperymentów prognostycznych na danych z wielu stacji pomiarowych, umożliwiających ocenę stabilności modeli oraz ich zdolności do generalizacji pomiędzy rozpatrywanymi lokalizacjami;

(3) Porównanie skuteczności zaproponowanych metod z metodami referencyjnymi oraz ocenę jakości prognoz z wykorzystaniem standardowych miar błędów predykcji umożliwiające ilościową weryfikację efektywności opracowanych modeli;

(4) Rozwinięcie eksperymentów prognostycznych na lata 2023 i 2024, pozwalające na stwierdzenie stabilności prognoz proponowanych metod.

Opracowane modele mogą wspierać proces podejmowania decyzji związanych z monitorowaniem stanu atmosfery oraz ograniczaniem skutków epizodów wysokiego zanieczyszczenia. Wiarygodne prognozy stężeń mogą być wykorzystane przez jednostki odpowiedzialne za zarządzanie środowiskiem do wcześniejszego identyfikowania okresów podwyższonego ryzyka wystąpienia epizodów wysokiego zanieczyszczenia powietrza, mogących prowadzić do niekorzystnych skutków zdrowotnych, planowania działań prewencyjnych oraz informowania społeczeństwa o spodziewanym pogorszeniu jakości powietrza. Tym samym uzyskane rezultaty przyczyniają się do rozwoju narzędzi wspomagających ochronę zdrowia publicznego oraz realizację działań ukierunkowanych na poprawę jakości środowiska.

Badania przeprowadzone w ramach niniejszej rozprawy doktorskiej, obejmujące analizę danych pomiarowych pochodzących z różnych lokalizacji w Polsce, pozwalają ocenić skuteczność opracowanych modeli w zróżnicowanych warunkach środowiskowych i emisyjnych. Jednocześnie przeprowadzona analiza wskazuje na dalsze możliwości rozwoju badań, związane przede wszystkim z integracją bardziej heterogenicznych źródeł danych, obejmujących m.in. sieci czujników niskokosztowych, stacje mobilne, systemy sensorowe oraz dane satelitarne. Takie podejście mogłoby zwiększyć pokrycie przestrzenne analiz i umożliwić prognozowanie jakości powietrza również na obszarach słabiej objętych monitoringiem. Istotnym kierunkiem pozostaje także rozwijanie modeli uwzględniających zależności czasowo-przestrzenne, integracja prognoz krótkoterminowych z analizami długoterminowymi i scenariuszowymi, a także szersze wykorzystanie metod interpretowalnej sztucznej inteligencji, pozwalających lepiej rozpoznawać wpływ czynników meteorologicznych, sezonowych i emisyjnych na wyniki modeli. Z punktu widzenia zastosowań praktycznych szczególne znaczenie ma dalsze powiązanie metod prognostycznych z systemami wczesnego ostrzegania oraz narzędziami wspomagania decyzji w zarządzaniu jakością powietrza. Ważnym aspektem pozostaje również uwzględnienie problemu brakujących danych pomiarowych, charakterystycznego dla rzeczywistych, rozproszonych systemów monitoringu jakości powietrza, który może stanowić element oceny odporności, skalowalności oraz użyteczności opracowywanych metod.

Rozważając zastosowanie opracowanych modeli, należy wskazać również pewne ograniczenia przeprowadzonych badań. Uzyskane wyniki są zależne od charakterystyki analizowanych zbiorów danych, w tym lokalizacji stacji pomiarowych, zakresu dostępnych zmiennych oraz specyfiki obserwowanych warunków atmosferycznych. Dodatkowym ograniczeniem jest niejednorodność danych wykorzystywanych w poszczególnych publikacjach, wynikająca z różnic w okresach pomiarowych, lokalizacjach oraz dostępnych parametrach wejściowych. Zakres przeprowadzonej walidacji obejmował określone zbiory danych i lokalne systemy monitoringu, dlatego konieczne są dalsze badania uwzględniające dane pochodzące z innych regionów, krajów oraz odmiennych typów sieci pomiarowych. Ponadto, przed zastosowaniem opracowanych modeli w rzeczywistych systemach operacyjnych, warto rozważyć dodatkowe etapy walidacji oraz ocenę ich przydatności w warunkach ciągłej eksploatacji.

Dodatek A

Załączniki – teksty publikacji i oświadczenia

A.1 Artykuł A1 – Załącznik 1

sensors

Review

A Comprehensive Review of Data-Driven Techniques for Air Pollution Concentration Forecasting

Jaroslaw Bernacki [1,*] and Rafał Scherer [2,3]

[1] Department of Computer Science and Engineering Systems, Wrocław University of Science and Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
[2] Department of Artificial Intelligence, Częstochowa University of Technology, al. Armii Krajowej 36, 42-200 Częstochowa, Poland; rafal.scherer@pcz.pl
[3] Faculty of Computer Science, AGH University of Kraków, 30-059 Kraków, Poland
* Correspondence: jaroslaw.bernacki@pcz.pl

Abstract

Air quality is crucial for public health and the environment, which makes it important to both monitor and forecast the level of pollution. Polluted air, containing harmful substances such as particulate matter, nitrogen oxides, or ozone, can lead to serious respiratory and circulatory diseases, especially in people at risk. Air quality forecasting allows for early warning of smog episodes and taking actions to reduce pollutant emissions. In this article, we review air pollutant concentration forecasting methods, analyzing both classical statistical approaches and modern techniques based on artificial intelligence, including deep models, neural networks, and machine learning, as well as advanced sensing technologies. This work aims to present the current state of research and identify the most promising directions of development in air quality modeling, which can contribute to more effective health and environmental protection. According to the reviewed literature, deep learning–based models, particularly hybrid and attention-driven architectures, emerge as the most promising approaches, while persistent challenges such as data quality, interpretability, and integration of heterogeneous sensing systems define the open issues for future research.

Keywords: air pollution; forecasting; machine learning; deep models; particulate matter

check for updates

Academic Editors: David Lopez-Carr and Kai Xu

Received: 14 August 2025
Revised: 17 September 2025
Accepted: 26 September 2025
Published: 1 October 2025

Citation: Bernacki, J.; Scherer, R. A Comprehensive Review of Data-Driven Techniques for Air Pollution Concentration Forecasting. *Sensors* **2025**, *25*, 6044. https://doi.org/10.3390/s25196044

1. Introduction

Air quality is crucial for human health and the functioning of ecosystems. Polluted air contains harmful substances such as particulate matter (PM), nitrogen oxides, or benzo(a)pyrene, which can cause serious respiratory and circulatory system diseases [1]. Long-term exposure to pollution increases the risk of asthma, chronic obstructive pulmonary disease (COPD), and cancer [2,3]. Children, the elderly, and people suffering from chronic diseases are particularly at risk [4]. Monitoring air quality allows for early warning of smog and taking action to reduce the emission of harmful substances. Thanks to appropriate regulations and initiatives, such as low-emission zones or the development of electric transport, air quality can be effectively improved. Taking care of clean air is an investment in the health of society and environmental protection [5].

The monitoring of air quality includes predicting the concentrations of various pollutants that have a significant impact on human health and the environment. The most frequently forecasted are particulate matter PM_{10} and $PM_{2.5}$, which, due to their small diameter, can penetrate the respiratory and circulatory systems, causing serious health problems [6]. Gases such as sulfur dioxide (SO_2), carbon monoxide (CO), and nitrogen oxides (NO and NO_2), which mainly come from the combustion of fossil fuels and road traffic, are also an important element of forecasts [7]. Particular attention is also paid to ozone (O_3), which is formed as a result of chemical reactions in the atmosphere and can be harmful at high concentrations. Depending on local meteorological conditions and emission sources, forecasts may also include other substances, such as benzo(a)pyrene or volatile organic compounds (VOCs). Indicators such as the Air Quality Index (AQI) [8] and the Air Pollution Index (API) [9] are used to classify air quality and inform the public about the level of risk. AQI used worldwide takes into account various pollutants and presents the results on an easy-to-read numerical and color scale, while API is an older indicator, used mainly in some Asian countries. There are also other metrics, such as AQHI [10], CAQI [11], DAQI [12], CAI [13], PSI [8], IAQI [14], which are used in different countries. A summary of the air quality indicators is presented in Table 1.

Table 1. Selected air quality indices adopted in different countries and regions.

Index	Country/Region	Main Pollutants Considered	Scale	Remarks
AQI	USA, China, India, others	$PM_{2.5}$, PM_{10}, O_3, CO, NO_2, SO_2	0–500	Widely used worldwide; health categories from "Good" to "Hazardous"
API	Malaysia (legacy), Hong Kong (before 2013)	PM_{10}, SO_2, NO_2, O_3, CO	0–500	Replaced by AQHI in Hong Kong in 2013
AQHI	Canada, Hong Kong (since 2013)	$PM_{2.5}$, O_3, NO_2	1–10+	Focused on health risk communication
CAQI	European Union (EEA)	$PM_{2.5}$, PM_{10}, NO_2, O_3, SO_2	0–100	Used in Copernicus/EEA air quality maps
DAQI	United Kingdom	$PM_{2.5}$, PM_{10}, O_3, NO_2, SO_2	1–10	Public information, combined with health advice
CAI	South Korea	$PM_{2.5}$, PM_{10}, O_3, CO, NO_2, SO_2	0–500	Includes both current and forecasted conditions
PSI	Singapore	$PM_{2.5}$, PM_{10}, O_3, CO, NO_2, SO_2	0–500	Based on US AQI but adapted to local standards
IAQI	China	Each pollutant separately (PM, gases)	0–500	Used as components of overall AQI

Information on predicted air quality also helps in planning environmental policy, e.g., limiting car traffic or introducing smog alerts. Modern methods based on artificial intelligence and numerical models allow for increasingly accurate analysis and forecasting of changes in air quality. The better the forecasts, the more effective preventive measures, which translates into improved public health and environmental protection. Therefore, the development and improvement of air quality forecasting methods should be one of the priorities in environmental policy [15]. The major air pollutants and their short descriptions are described in Figure 1.

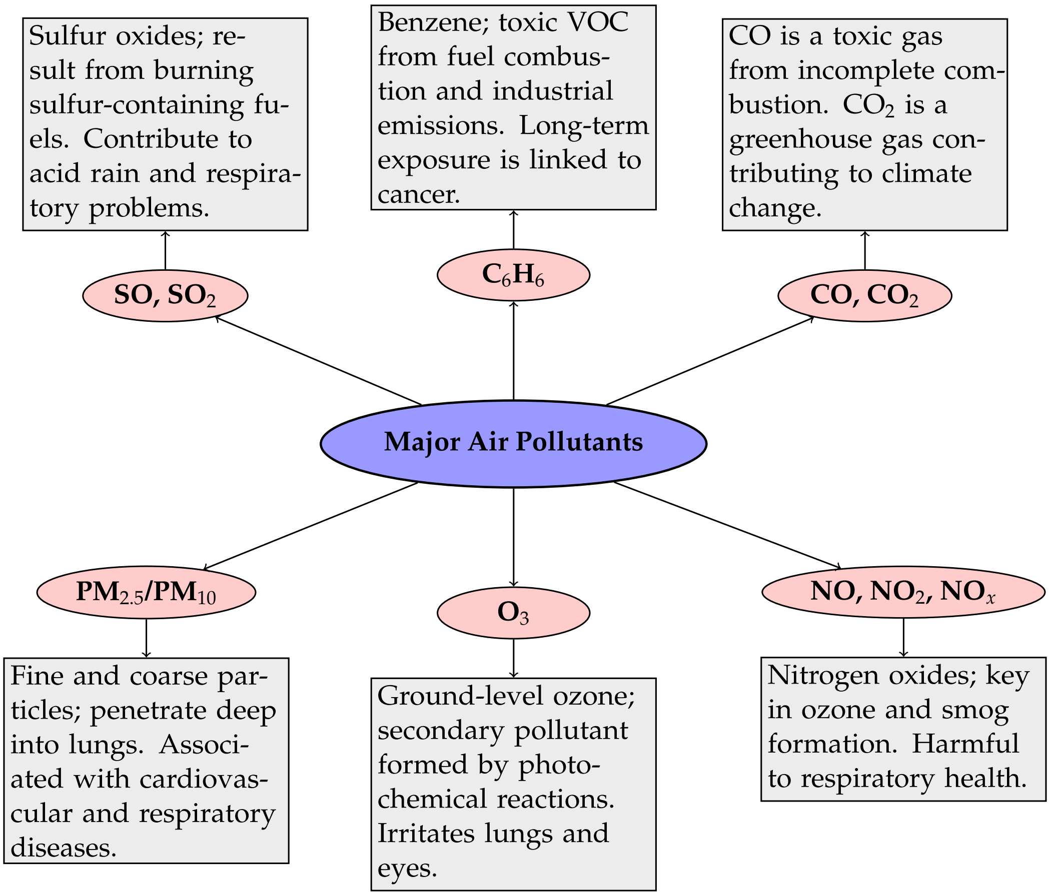


Figure 1. Major air pollutants and their short descriptions [16].

In this paper, we review air quality forecasting methods, including modern techniques based on artificial intelligence, as well as classical approaches based on statistical foundations. In particular, we analyze deep models, such as neural networks, which can effectively predict pollution levels. We also present machine learning methods that use historical data and environmental factors to forecast air quality in various conditions. Additionally, we discuss statistical approaches that are still used in trend analysis and short-term forecasting. Moreover, we analyze advanced sensing technologies, including low-cost, mobile, and IoT-based air quality monitoring systems, which enable flexible and scalable data acquisition. These sensor-based solutions, when properly calibrated and enhanced with AI, can complement traditional monitoring networks and significantly improve data granularity. We analyze the efficacy of these methods, indicating their advantages and limitations in the context of different forecasting scenarios. Therefore, the goal of the paper is to present the current state of research and indicate the most promising directions of development in the field of air quality modeling and monitoring. The relations of forecasting methods discussed in the paper are presented in Figure 2.

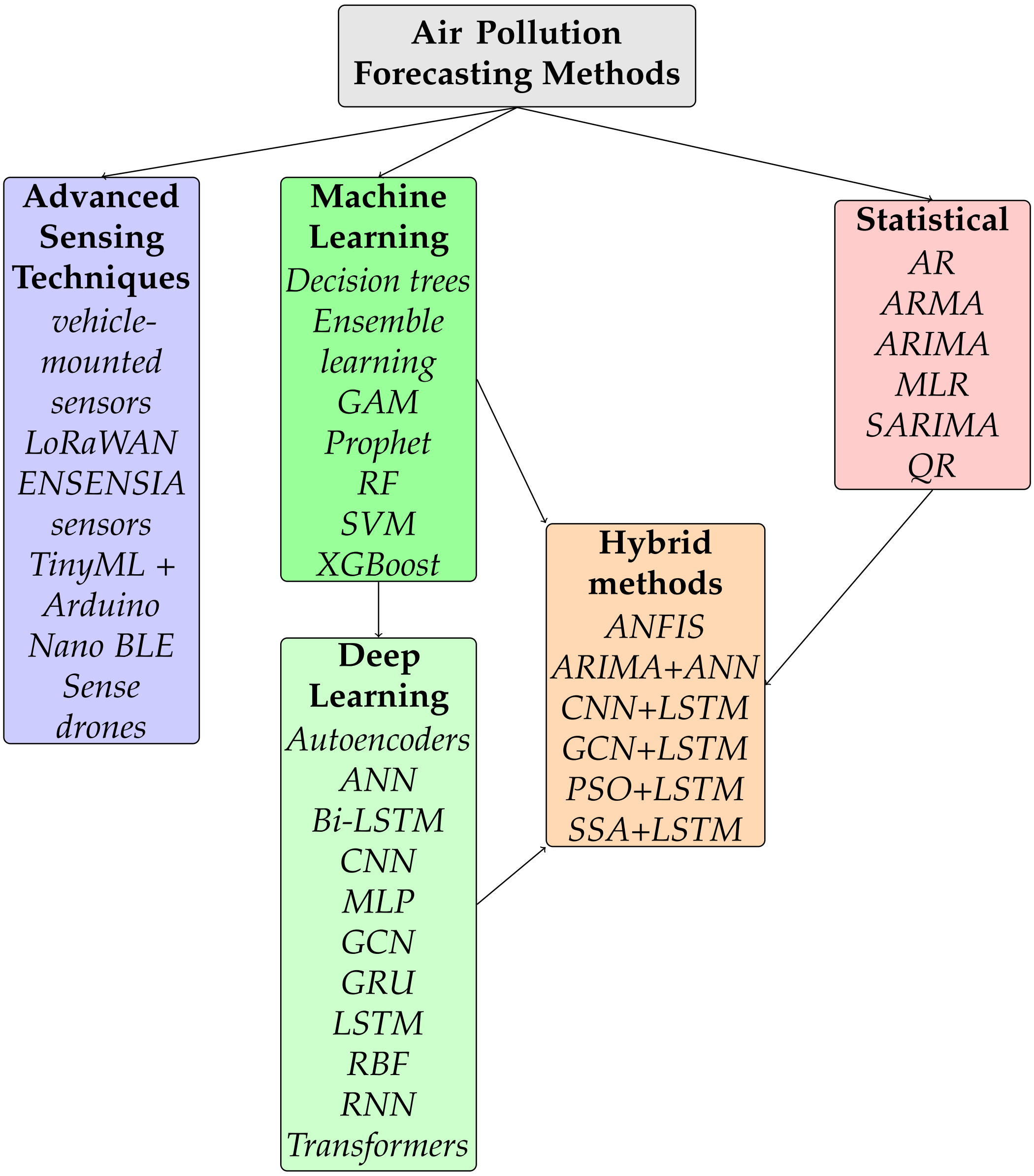


Figure 2. Categorization of forecasting methods with selected algorithms/methods. Deep learning is a subcategory of machine learning [16].

Organization of the paper and conventions

The paper adopts the structure of a traditional narrative review and is organized as follows. In Section 2, we describe the advanced sensing techniques for air pollution monitoring and detection. Section 3 describes neural networks and deep learning models. In Section 4, the description of classical machine learning methods is included. In the Section 5, the statistical methods for forecasting are described. Section 6 presents the hybrid methods. Each section is summarized in a table and a discussion subsection, which contains all the methods discussed. In Section 7, other methods for forecasting are considered. Section 8 summarizes the paper, with the directions of open issues and future work.

At the end of the paper, there is an appendix that presents full explanations of the abbreviations of the forecasting methods reviewed in this survey. Additionally, it includes formulas for all quality measures used in the analysis, which facilitates their interpretation and a comparison of results.

2. Advanced Sensing Techniques

This section presents an overview of air pollution detection methods using various types of sensors, both stationary and mobile. Technologies based on low-cost sensors, drones, and wireless sensor networks are discussed. The advantages and limitations of each solution are also highlighted in terms of measurement accuracy, weather resistance, and integration with modern data processing technologies.

In [17], a low-cost LoRaWAN wireless network for particulate matter (PM) measurements, tested in an industrial setting (Riotinto mine), is developed. The devices are based on SDS011 sensors and Arduino and Raspberry Pi components. Local calibration was performed, achieving an MAE of less than 0.3 $\mu g/m^3$ and a coefficient of determination of $R^2 = 0.96$. The network enables real-time data collection and transmission to a cloud platform. The authors point to the solution's high cost-effectiveness for environmental monitoring.

The aim of the study [18] was to calibrate a low-cost NO_2, NO, O_3, CO, and $PM_{2.5}$ sensors called ENSENSIA in an urban environment using machine- and deep learning models. Data collected in a German city were compared with official reference stations. After calibration, high agreement was achieved ($R^2 > 0.86$) and measurement error was reduced. The authors demonstrate that onsite calibration significantly improves the reliability of measurements. The study emphasizes the need for dynamic correction of low-end sensors.

The paper [19] describes the design of a low-cost drone-mounted air quality measurement system for the city of Opole, Poland. The system utilizes a PMS7003 sensor for $PM_{2.5}$ and PM_{10} detection, MQ sensors for gas detection (NO_2, CO), and a GPS module. The total cost of the device was under EUR 70, making it extremely competitive. The results showed that the UAV-derived data correlated with ground-based data. The paper demonstrates the potential of drones for the rapid and cost-effective identification of smog-prone areas.

In [20], the drones equipped with laser sensors for detecting gaseous pollutants (C_6H_6, HCHO, SO_2); and particulate matter (PM_1, $PM_{2.5}$, and PM_{10}) and GPS sensors to monitor linear emissions (e.g., along roads and railways) were used. The devices analyzed pollutants in the vertical profile of the atmosphere and along transportation routes. The drones operated autonomously and collected data under variable atmospheric conditions. The results were compared with data from reference stations. This work suggests that UAVs are the future of microregional emission source detection.

In [21], the Enviro-IoT system, which is a low-cost, real-time air quality monitoring solution based on miniature sensors and IoT technology, is presented. The authors describe the four-year design and implementation process of the system, then present the results of a nine-month study in urban environments. The system was compared with industrial-grade testing equipment and achieved high measurement accuracy for $PM_{2.5}$, PM_{10}, and NO_2 (98%, 97%, and 97%, respectively). The analysis included over 57,000 samples, confirming the solution's effectiveness. The paper demonstrates that low-cost sensors combined with IoT can be an effective alternative to traditional measurement systems.

The research [22] involves developing a low-cost, self-built air quality assessment system that can be used both indoors and outdoors. The system is based on ESP32 microcontrollers using the ESP-NOW wireless communication protocol and Alphasense sensors to measure the concentration of aerosols and gases such as NO_2, SO_2, O_3, CO, CO_2, and H_2S. The device was used to assess air contaminated with tobacco smoke after passing through a DBD plasma purification reactor. A significant reduction in aerosol and SO_2 particles was observed, while an undesirable increase in NO_2 was observed. The measurement results were consistent with trends observed with the professional GRIMM device, confirming the potential of the developed system as a cost-effective and effective alternative to air quality monitoring.

The work [23] focuses on developing a universal method for calibrating low-cost air quality sensors, enabling their mass deployment in urban monitoring networks. Standard approaches require individual calibration of each sensor, which generates high costs and limits scalability. The authors propose a global calibration method without transfer samples, based on data from a limited number of sensors and machine learning algorithms, that can be applied to all devices of a given type. Tests conducted at different times of the year have shown that the effectiveness of this method is comparable to approaches requiring individual calibration parameters. Implementing such calibration could significantly reduce costs and enable the widespread use of accurate, low-cost IoT devices for air quality monitoring.

The study [24], conducted in Delhi, examined three popular PM sensor models (SPS30, PMS7003, HPMA115C0-004) in high humidity and dust conditions for $PM_{2.5}$ detection. The work highlights the need for dynamic calibration and the implementation of real-time humidity compensation. Long-term sensor drift analysis is also conducted.

In [25], the application of TinyML platforms to real-time ozone prediction using low-cost devices such as the Arduino Nano BLE Sense is demonstrated. The aim of the system is to detect the CO concentrations, using temperature and air pressure. The solution is low-cost and operates in real time, and can be mounted on drones.

In [26], the calibration of low-cost particulate matter sensors, which are gaining popularity due to their low price and portability, is addressed. Standard calibration methods require large amounts of data from simultaneous reference measurements, which is time-consuming and expensive. The authors propose the use of transfer learning methods, which allow for efficient sensor calibration with minimal interaction time with the reference device. In particular, they present a meta-learning approach that utilizes data from other sensors and a limited amount of data from the target sensor. Experiments on $PM_{2.5}$ concentrations show that this method outperforms other compared approaches in terms of calibration efficiency.

The research presented in [27] proposes the creation of high-resolution $PM_{2.5}$ maps by combining data from 320 mobile vehicle-mounted sensors and 52 stationary stations. Machine learning algorithms were used for data correction and spatial interpolation. The maps achieved a resolution of 500 m and a 5-min refresh rate, with a global error of +4.3%. The work demonstrates the enormous potential of hybrid (mobile-stationary) systems for managing urban air quality.

The paper [28] presents a low-cost, multi-parameter air quality monitoring system that utilizes various machine learning algorithms for calibration. The system simultaneously measures the concentration of particulate matter ($PM_{2.5}$, PM_{10}) and gases (SO_2, NO_2, CO, O_3), taking into account temperature and humidity. A prediction model was developed and compared with data from reference devices, evaluating the performance of different ML algorithms. The results demonstrate high prediction accuracy (R^2 up to 0.99) with low RMSE and MAE errors. It was demonstrated that the system supported by ML algorithms can effectively predict the concentration of particulate matter and gases in the air.

The paper [29] presents a low-cost air quality monitoring platform based on an unmanned aerial vehicle (UAV) equipped with specialized sensors for measuring CO, O_3, and NO_2. The solution aims to increase the mobility and availability of measurements in hard-to-reach locations, addressing the limitations of traditional measurement networks, such as high costs and a lack of mobility. The system enables real-time data collection and UAV tracking using LoRa communication and the PTECA web application. The platform provides data visualization and live monitoring during sample collection. Initial tests in controlled airspace using geofencing confirmed the effectiveness and practicality of the proposed solution.

The advanced sensing techniques are summarized in Table 2.

Table 2. Advanced sensing techniques for air pollutants forecasting.

Reference	Air Pollutants	Country/Region	Sensors/Techniques
Zafra et al. [17]	PM_{10}	Europe/Spain	LoRaWAN
Apostolopoulos et al. [18]	CO, NO, NO_2, O_3, $PM_{2.5}$	Europe/Greece	ENSENSIA sensors
Pochwała et al. [19]	CO, NO_2, PM_{10}, $PM_{2.5}$	Europe/Poland	utilizing drones
Jaron et al. [20]	C_6H_6, HCHO, SO_2, PM_1, PM_{10}, $PM_{2.5}$	Europe/Poland	drones
Johnson et al. [21]	NO_2, PM_{10}, $PM_{2.5}$	Europe/England	Enviro-IoT sensors system
Moskal et al. [22]	CO, CO_2, H_2S, NO_2, O_3, SO_2	Europe/Germany	ESP32 boards + wireless communication protocol ESP-NOW
De Vito et al. [23]	$PM_{2.5}$	Europe/Norway	Low Cost Air Quality Monitoring System
Kumar et al. [24]	$PM_{2.5}$	Asia/India	SPS30, PMS7003, HPMA115C0-004 sensors
Ken et al. [25]	CO	Asia/India	TinyML + Arduino Nano BLE Sense
Yadav et al. [26]	$PM_{2.5}$	Asia/India	low-cost particulate matter sensors
Xu et al. [27]	$PM_{2.5}$	Asia/Hong Kong	vehicle-mounted sensors
Wang et al. [28]	CO, NO_2, O_3, SO_2, PM_{10}, $PM_{2.5}$	Asia/China	low-cost, multi-parameter air quality monitoring system
Medina et al. [29]	CO, NO_2, O_3	Sth. America/Colombia	low-cost air quality monitoring platform + drones + LoRA + PTECA

Discussion

Recent advances in low-cost air quality monitoring systems demonstrate remarkable diversity in sensing platforms, sensor types, and data processing techniques. Many studies emphasize mobile and flexible deployment, such as drone-based platforms and vehicle-mounted sensors, which offer high spatial resolution in dynamic environments. Fixed IoT-based networks like Enviro-IoT or LoRaWAN provide real-time coverage at low cost. A common theme is the need for calibration, either via on-site methods or via scalable machine learning solutions. Overall, modern systems combine compact sensors with machine learning methods, wireless communication (e.g., ESP-NOW, LoRa), and sometimes TinyML, enabling efficient and accurate air pollution detection across various environments.

In addition to low-cost and mobile sensing platforms, a wide range of sensor types is employed in air quality monitoring, each offering distinct advantages and limitations. IoT-enabled low-cost devices provide high-density local coverage, while reference-grade ground stations remain the standard for accuracy and regulatory compliance. Satellite-based imaging systems contribute large-scale spatial coverage and long-term historical records, although with coarser temporal resolution. Aircraft- and drone-mounted sensors fill the gap by delivering flexible, high-resolution observations in areas that are otherwise inaccessible. Together, these complementary approaches highlight the importance of sensor diversity in environmental monitoring, as no single technology alone can provide the accuracy, coverage, and scalability required for comprehensive air quality forecasting.

However, despite these advances, low-cost sensors remain prone to drift, humidity sensitivity, and degradation over time, which can compromise data reliability; in this context, machine learning models are particularly valuable because they can dynamically learn correction patterns, compensate for non-linear sensor behavior, and extend the practical usability of such systems. Low-cost solutions provide significant flexibility, mobility, and scalability, making them attractive for dense urban deployments; however, their effectiveness is often constrained by challenges such as long-term stability, susceptibility to atmospheric conditions (e.g., humidity and temperature fluctuations), signal drift, and the limited lifetime of inexpensive components. Although local calibration procedures and ML algorithms can decrease some of these shortcomings, persistent issues related to standardization, interoperability, power quality, and data transmission limit their widespread adoption in industrial or institutional contexts.

By contrast, satellite-based and conventional ground monitoring stations offer higher measurement accuracy and broader spatial coverage but are associated with high costs, limited accessibility, and lower temporal resolution. These differences highlight the importance of data quality and diversity, as relying on a single type of monitoring source can lead to biased or incomplete forecasts. Future research directions should therefore focus on designing robust fusion strategies that integrate heterogeneous datasets from multiple sensor types, exploiting their complementary strengths to achieve improved forecasting accuracy, resilience to missing or noisy data, and broader applicability across different regions and time scales. In particular, hybrid frameworks that combine high-resolution local sensor data with the wide coverage of satellite observations may provide a more holistic representation of pollution dynamics and support more effective environmental decision-making. Moreover, integrating heterogeneous data sources, such as IoT-based networks, mobile monitoring platforms, and satellite observations, could further enhance model generalizability and robustness, ensuring that forecasting systems remain accurate across diverse environments and evolving pollution scenarios.

3. Neural Networks and Deep Learning Models

In this section, we present methods for forecasting air pollution concentrations based on deep learning. Modern neural networks, thanks to their ability to model complex nonlinear relationships, show high efficiency in predicting changes in air quality. We will discuss various network architectures, i.e., artificial neural networks (ANN), multilayer perceptron (MLP), convolutional neural networks (CNN), long short-term memory, recurrent neural networks (RNN), which are used in the analysis of time series of pollution.

3.1. Artificial Neural Networks

The paper [30] uses the artificial neural network method and the group data handling method to predict CO_2 emissions based on energy sources. The experiments used data from the following five countries: Iran, Kuwait, Qatar, Saudi Arabia, and the UAE. The evaluation by using the R^2 confirmed its efficacy in predicting CO_2 emissions.

In [31], an artificial neural network-based integrated model to predict the indoor environmental parameters, including predicted mean vote and concentrations of CO_2, PM_{10}, and $PM_{2.5}$, in a school building was proposed. The model was optimized using a multi-objective genetic algorithm and prevents spurious regression by excluding irrelevant input variables. The prediction accuracy was evaluated with root mean square error, which demonstrated high accuracy and applicability for building environment control and energy optimization.

The research [32] evaluates the ability of an artificial neural network to predict hourly air pollutant concentrations and air quality indices for Ahvaz (Iran), one of the most polluted cities due to dust storms. The proposed model, using meteorological data, pollutant concentrations, and time-related factors, showed a higher correlation coefficient (R) and a lower root mean square error (RMSE) for six criteria pollutants, compared with other models.

In the paper [33], neural networks combined with meteorological data were used to predict hourly $PM_{2.5}$ concentrations. The experiments were conducted using data from the Beijing Aotizhongxin monitoring station. The proposed method demonstrated high prediction accuracy, achieving better R^2 values compared to other approaches.

The study in [34] applied an artificial neural network to forecast $PM_{2.5}$ concentrations in Liaocheng. The model, consisting of 11 input parameters, a hidden layer with 6 neurons, and a single output, was trained on 80% of the dataset and validated on 10%. The analysis revealed a gradual decline in $PM_{2.5}$ levels, with the Bayesian Regularization algorithm providing the most accurate predictions and yielding high correlation coefficients. These results demonstrate the ability of ANN to capture nonlinear relationships and to effectively predict monthly $PM_{2.5}$ concentrations. In addition, [35] reported the use of ANN for forecasting hourly $PM_{2.5}$ and PM_{10} levels in Chongqing.

The study in [36] explored the application of artificial neural networks (ANN) and wavelet transform neural networks (WANN) for forecasting daily $PM_{2.5}$ concentrations in Shanghai. Between 2014 and 2020, $PM_{2.5}$ levels declined by 39.3%. The models utilized meteorological variables and pollutant concentrations from previous days, with minimum and maximum temperature and maximum atmospheric pressure identified as the most influential factors. Among the tested models, WANN with the tansig–purelin activation function consistently outperformed ANN at all stages of evaluation, demonstrating high predictive accuracy and confirming its suitability for $PM_{2.5}$ forecasting in support of air quality management strategies.

In [37], an artificial neural network model using meteorological and PM data to predict daily concentrations of $PM_{2.5}$ and PM_{10} is considered. Among the four tested models, the wavelet-based multilayer perceptron neural network outperformed the others, especially when utilizing lagged wind speed as a predictor. The proposed model achieved high accuracy, with R^2 values up to 0.91 for $PM_{2.5}$ and 0.98 for PM_{10}, and low RMSE values. These findings demonstrate the model's potential for reliable air quality forecasting and issuing timely health alerts.

In [38], an artificial neural network model to forecast short-term CO_2 emissions in the cereal sector of Apulia (Southern Italy) is proposed. The model aims to evaluate the environmental impact of adopting sustainable farming methods on pollution reduction. The approach offers a flexible alternative to classical methods, which are often too rigid for modeling complex environmental phenomena. The goal is to explore how more eco-friendly agricultural practices can help reduce CO_2 emissions in the region.

The research in [39] explores machine learning-based prediction of nitrogen dioxide (NO_2) concentrations, utilizing historical pollutant levels and meteorological data. Focusing on Erfurt (Germany), the study uses temporal dependencies through embedded variables, enabling the model to account for local factors such as traffic and special events. Additionally, the approach utilizes pollutant seasonality and seven meteorological parameters to predict NO_2 levels in the upcoming hours. Experimental results demonstrate higher forecasting accuracy compared to other models, particularly during disruptions in typical seasonal patterns, such as holidays. These predictions can support emission management strategies, including traffic regulation.

3.2. Convolutional Neural Networks

In [40], convolutional neural network models were employed to predict real-time PM_{10} and $PM_{2.5}$ concentrations in Poland. The evaluation, based on R^2 and RMSE measures, confirmed the models' high accuracy in forecasting the analyzed concentrations.

The paper [41] presents a deep learning model for forecasting $PM_{2.5}$ air quality, addressing the challenges posed by the nonlinear and dynamic nature of multivariate air quality time series data. The proposed model integrates one-dimensional convolutional neural networks (1D-CNNs) to capture local trends and spatial correlations, along with bi-directional long short-term memory networks (Bi-LSTM) to learn spatial–temporal dependencies. This architecture is designed to extract shared features from multivariate time series data related to air quality. Extensive experiments with real-world datasets demonstrate that the model achieves high accuracy in $PM_{2.5}$ forecasting, making it a promising approach for air pollution prediction and management.

In [42], the prediction of particulate matter concentrations in Beijing (China), using deep learning methods, is considered. The paper utilizes a convolutional neural network and a long short-term memory network model to predict hourly PM concentrations, utilizing meteorological data, historical pollutant data, and concentrations from nearby stations as spatial–temporal features. The performance of various deep learning algorithms, including LSTM, Bi-LSTM, GRU, Bi-GRU, and CNN, was compared, with the CNN-LSTM model outperforming all others in predictive accuracy. These results demonstrate the effectiveness of combining convolutional neural networks and long short-term memory networks for air quality prediction.

The work [43] developed a model combining convolutional neural networks and long short-term memory networks for air pollution prediction using sensor data. It considered both univariate and multivariate models, with the latter using data from multiple pollutants and meteorological factors. The model was tested using data from Barcelona, Kocaeli, and Istanbul, showing significant improvements in prediction accuracy compared to traditional LSTM models. Transfer learning was also applied, with better results when transferring the model from Kocaeli to Istanbul, highlighting the model's adaptability to different cities.

The study [44] proposed a spatiotemporal convolutional neural network and long short-term memory model to predict daily average $PM_{2.5}$ concentrations in Beijing City (China). The model used both linear and nonlinear correlations between target and observation parameters, using historical air quality and meteorological data from 384 monitoring stations across China. By efficiently extracting features through CNN and reflecting long-term data trends with LSTM, the model achieved accurate and fast predictions. Compared to multilayer perceptron and LSTM models, the CNN-LSTM method demonstrated better stability and performance in air quality prediction.

The study presented in [45] introduces a deep learning model that combines a convolutional neural network with a long short-term memory network to predict hourly total suspended particle (TSP) concentrations. The CNN serves as a data processor, utilizing feature extractors based on statistically significant lagged variables. The LSTM network utilizes a new feature mapping scheme to predict the next hourly TSP concentration. Extensive testing of the proposed model demonstrates its better performance compared to an ensemble of five other machine learning models.

In [46], a CNN-LSTM model is developed to forecast the $PM_{2.5}$ concentration for the next 24 h in Beijing. This model combines the strengths of both convolutional neural networks and long short-term memory networks: CNN efficiently extracts air quality-related features, while LSTM captures the long-term temporal context of the data. The model takes air quality data from the past 7 days as input and predicts the next day's $PM_{2.5}$ concentration. The study compares four models: univariate LSTM, multivariate LSTM, univariate CNN-LSTM, and multivariate CNN-LSTM. The results indicate that the proposed multivariate CNN-LSTM model outperforms the others, achieving the best performance with low error and short training time.

3.3. Attention-Based Methods

In the work [47], the authors propose a model combining a wind-sensitive attention mechanism, an LSTM network, and XGBoost to predict $PM_{2.5}$ concentrations, taking into account the influence of wind direction and speed on pollutant transport. The input data included $PM_{2.5}$ measurements from neighboring areas and meteorological forecasts. Experiments were conducted using real $PM_{2.5}$ data and weather forecasts.

The research [48] proposes a new air quality forecasting model based on the IMDA-VAE architecture, which combines a variational autoencoder and an attention mechanism. The method is experimentally validated using pollution data from four US states. Evaluation based on six statistical metrics demonstrates that the model effectively predicts concentrations of various pollutants at multiple locations. The results show that IMDA-VAE outperforms existing models, including LSTM, GRU, and their attention-based versions.

The authors of [49] developed a new air quality forecasting model based on the Bi-LSTM convolutional autoencoder with attention. The study was conducted in Busan (South Korea), using data on particulate matter concentration, meteorological conditions, and road traffic volume obtained using the YOLO algorithm. The model effectively captured spatial–temporal relationships and significantly outperformed previous approaches in both short- and long-term forecasts. Furthermore, it was successfully deployed in the cloud, enabling real-time air quality forecasting.

In [50], an advanced Seq2Seq model, called the attention-based air quality forecasting model (ABAFM), which replaces the RNN encoder with a pure attention mechanism with positional embedding is proposed. This modification reduces training time and enhances the robustness of the model. Experiments conducted at Olympic Center and Dongsi monitoring stations in Beijing show that ABAFM outperforms existing methods, particularly in predicting sudden changes in $PM_{2.5}$ levels.

In the study [51], an image-based deep learning model combining a 3D-CNN, GRU, and attention mechanism to forecast air quality is proposed. The 3D-CNN extracts spatial features from images, the GRU captures temporal dependencies, and the attention mechanism adjusts feature influence to reduce fluctuations. Tested on the Shanghai scenery dataset with air quality data, the model outperformed other state-of-the-art methods. It provides accurate multi-horizon predictions and reliable early warnings for air pollutants.

3.4. Multilayer Perceptron

The paper [52] examines the forecasting performance of multiple models for predicting PM_{10} concentrations on the East Coast of Peninsular Malaysia, where PM_{10} is a significant air quality concern. The research compares a linear model (multiple linear regression) with two non-linear models (multilayer perceptron and radial basis function) using meteorological and pollutant historical data at urban, suburban, and rural sites. The non-linear radial basis function model outperforms the linear model, reducing forecasting errors by 78.9% in urban areas, 32.1% in suburban areas, and 39.8% in rural areas. The results highlight the complexity of the relationship between PM_{10} and its contributing factors, with the non-linear models, particularly artificial neural networks, offering more accurate forecasts. This model provides reliable next-day PM_{10} concentration predictions, useful for early warning purposes.

In the work [53], forecasting maximum O_3 concentrations in urban microlocations is addressed. It proposes an integrated modeling approach using multilayer perceptron neural networks, which combine data from the QualeAria air quality model, WRF weather prediction model, and local air pollution measurements. This integrated approach improves 1-day-ahead ozone forecasts, particularly in areas where local resolution is lacking. The results show that combining these models enhances forecast accuracy and provides better alert systems for exceeding ozone thresholds in Slovenia.

In the study [54], a comparison of the performance of multiple linear regression and multilayer perceptron in predicting the SO_2 concentration in Tehran's air is conducted. Different parameters, such as meteorological conditions, urban traffic data, green areas, and time variables, were used for the analysis. The results show that the MLP model achieved better performance in terms of R^2 and RMSE measures than the MLR. Sensitivity analysis showed that the most important factors affecting the SO_2 level are the one-day time lag, park index, season, and total park area. The MLP model can support the analysis and air quality management strategies, emphasizing the importance of artificial neural networks in reducing urban pollution.

3.5. Other DL Approaches

In [55], recurrent artificial neural networks to estimate SO_2 and PM_{10} air pollution levels in Sakarya (Turkey) were proposed. The results showed strong correlations and low RMSE values, demonstrating the efficacy of deep learning methods even under challenging conditions. The study also highlighted that PM_{10} levels in Sakarya often exceeded legal limits, particularly during peak industrial periods.

In [56], the authors proposed an aggregated long short-term memory model for air pollution forecasting. Its performance was evaluated against three commonly used machine learning approaches, using MAE, RMSE, and MAPE as assessment metrics. The experiments were conducted with air quality data collected from multiple monitoring sites across Taiwan.

In [57], a radial basis neural network was employed for air quality prediction in Zhengzhou and Shanghai (China). The study explores the network's ability to model and forecast air quality based on the collected data from these cities.

The study [58] uses an artificial neural network and wavelet ANN to forecast the air pollution index (API) by analyzing its linear and nonlinear relationships with meteorological variables in Xi'an and Lanzhou. The study identifies optimal input variables for forecasting, which include APIs from the previous days and selected meteorological factors. The WANN model, utilizing Bayesian regularization, outperformed the standard ANN model in both cities, demonstrating its ability to accurately predict API values by capturing nonlinear relationships. The findings suggest that WANNs are effective for short-term API forecasting, potentially aiding in environmental management policies.

The research presented in [59] introduces an artificial neural network-based approach using a feed-forward neural network to predict NO_x emissions from diesel-powered underground LHD vehicles. The model achieves predictions with less than 15% error, making it an efficient tool for estimating dynamic behavior in complex systems. The accuracy enables its application in estimating environmental impact and planning ventilation systems in mines. This approach offers insights into underground air quality and enhances safety for mining crews.

The paper [60] considers a factor-augmented autoregressive neural network to forecast NO_x pollution levels in Madrid, using lagged NO_x values, meteorological data, and latent factors. The model outperforms competing methods, providing more accurate predictions of air pollution, which could aid policymakers in improving air quality monitoring and management.

The study [61] develops a combined weight forecasting model for predicting NO_2 concentration in Beijing using wavelet-transformed data. Three models using the discrete wavelet transform with long short-term memory, gated recurrent unit, and bi-directional long short-term memory are constructed and compared. The proposed model integrates these models through weighted assignment, utilizing their strengths to enhance prediction accuracy. Results indicate that it outperforms individual models, making it a more effective tool for air quality forecasting and governance efforts.

In [62], a graph convolutional temporal sliding long short-term memory model to predict air pollutant concentrations by capturing spatiotemporal correlations is proposed. The model integrates graph convolutional networks for spatial dependency modeling and long short-term memory networks with a temporal sliding strategy for temporal dependency modeling. Applied to the Beijing–Tianjin–Hebei region, it predicts $PM_{2.5}$, PM_{10}, O_3, CO, SO_2, and NO_2 levels for 24 h, outperforming state-of-the-art models in accuracy and stability. The approach supports improved air quality management and decision-making.

The research [63] proposes a DL-Air model, which is a hierarchical deep learning model for air quality forecasting. DL-Air consists of three components: an encoder that captures spatial relationships in air quality data, a variant of LSTM that models temporal dependencies and spatiotemporal correlations, and a decoder that transforms latent predictions into actual forecasts. Evaluated on real-world data from Delhi, DL-Air outperformed baseline models, significantly reducing RMSE and MAE and improving R^2, while maintaining consistent predictive performance across all seasons.

The study [64] implements a transformer-based model, the temporal fusion transformer, trained on observational data from three European countries, achieving high accuracy in four-day forecasts of a daily maximum 8 h of ozone. The model outperforms other machine learning approaches and generalizes well to data from 13 additional European countries not used in training. Analysis of the model's attention mechanism shows it focuses on key physical variables and the most relevant past ozone values, highlighting its potential as an efficient tool for air quality forecasting across Europe.

In the work [65], a lightweight deep learning model based on sparse attention transformer networks (STN) with encoder–decoder layers, which reduces time complexity while learning long-term dependencies from $PM_{2.5}$ data, is proposed. Experiments on the Beijing and Taizhou datasets demonstrate that the STN model outperforms state-of-the-art methods in both short-term and long-term predictions, achieving high R^2 and low RMSE and MAE values. The model remains effective for multi-step forecasts up to 48 h, highlighting its robustness and efficiency for air quality prediction.

The paper [66] presents HighAir, a hierarchical graph network (HGN) for air quality forecasting that utilizes an encoder–decoder architecture and considers factors such as weather and land use. The model combines city- and station-level analysis, employing message-passing mechanisms and specialized strategies for exchanging information between levels. Furthermore, edge weights are dynamically adjusted based on wind direction, allowing for a more realistic representation of pollutant dispersion. Tests conducted on data from the Yangtze River Delta (10 cities, 61,500 km^2) show that HighAir significantly outperforms existing air quality forecasting methods.

The authors of paper [67], propose an air quality forecasting method based on an attention temporal graph convolutional network (AT-GCN), which combines attention, a recurrent GRU network, and graph convolutional mechanisms. The model was tested on air quality, meteorological, and road traffic data from Madrid. Results compared with other methods (LSTM, GRU) showed that the proposed approach better captures complex dependencies and achieves lower prediction errors. Research indicates the significant potential of AT-GCN for urban pollution prediction applications.

In [68], a spatiotemporal graph network model based on GraphSAGE for forecasting ozone concentrations in urban areas is proposed. The model was trained and tested on data from Houston, Texas, USA, considering different forecast horizons (1, 3, and 6 h), the number of time lags, and data from neighboring stations. The results showed significant improvements over the persistence model—by 33.7%, 48.7%, and 57.1%, respectively—and lower errors compared to the existing literature. SHAP analysis revealed that solar radiation becomes more important at longer forecast horizons, and the model correctly identified exceedances of EPA air quality standards.

The research [69] proposes a group-aware graph neural network (GAGNN), which is a hierarchical model for national-scale air quality forecasting. The model constructs a city graph and a city-group graph to capture both spatial dependencies and hidden connections between distant but correlated locations. A differentiable clustering network is introduced to discover inter-city dependencies and a module for encoding inter-group correlations, and a message-passing mechanism is then applied to the graph. Experiments conducted on large datasets from China and the United States show that GAGNN outperforms existing air quality forecasting models.

In [70], a computational framework based on graph neural networks for air quality prediction in complex urban environments is proposed. This model considers both the spatial distance between stations but also contextual factors such as land use and dominant pollutant sources. Additionally, an approach is developed to handle heterogeneous measurement systems, where stations monitor different sets of pollutants. The effectiveness of the solution was validated on a nationwide Spanish air quality dataset, achieving very promising results.

The paper [71] presents a self-supervised hierarchical graph neural network (SSH-GNN) method for air quality forecasting with high spatiotemporal resolution. The model first approximates the air quality distribution across a city based on historical data and contextual factors such as weather and traffic volume. It then uses a hierarchical recurrent graph network to model long-range spatiotemporal correlations between urban regions. Using spatio-temporal self-supervision strategies, SSH-GNN can effectively capture both universal topological and contextual patterns, resulting in significant improvements in forecast accuracy when tested on two real-world datasets.

The paper [72] presents a real-time air pollution forecasting system for five key locations in Delhi, using historical air quality and meteorological data. The authors propose an end-to-end sequential modeling framework to predict concentrations of pollutants such as NO_2, $PM_{2.5}$, and PM_{10}, and to classify their threat levels for the next 24 h. The system effectively handles missing or unreliable data and demonstrates higher performance compared with baseline forecasting methods.

A comprehensive survey of different techniques using deep learning for air pollution forecasting is presented in [16]. The paper reviews various approaches, including artificial neural networks. The study emphasizes the effectiveness of deep learning in improving the accuracy and reliability of air quality predictions, highlighting their ability to model complex, non-linear relationships between meteorological factors, pollutant concentrations, and other influencing variables. It also discusses the challenges of implementing these techniques in real-world air quality forecasting systems, such as data quality and computational cost.

The deep learning methods are summarized in Table 3.

Table 3. Deep learning methods for air pollutants forecasting.

Reference	Air Pollutants	Methods
Ahmadi et al. [30]	CO_2	
Cho et al. [31]	CO_2, PM_{10}, $PM_{2.5}$	
Sohn et al. [32]	CH_4, CO, NO, NO_2, O_3, SO_2	
Liang et al. [33]	$PM_{2.5}$	
He et al. [34]	$PM_{2.5}$	
Guo et al. [35]	PM_{10}, $PM_{2.5}$	ANN
Guo et al. [36]	$PM_{2.5}$	
Gogikar et al. [37]	PM_{10}, $PM_{2.5}$	
Gallo et al. [38]	CO_2	
Ramentol et al. [39]	NO_2	embeddings + ANN
Chae et al. [40]	PM_{10}, $PM_{2.5}$	CNN
Du et al. [41]	$PM_{2.5}$	CNN + Bi-LSTM
Bekkar et al. [42]	PM	
Gilik et al. [43]	NO_2, NO_x, O_3, PM_{10}, SO_2	
Pak et al. [44]	$PM_{2.5}$	CNN-LSTM
Sharma et al. [45]	TSP	
Li et al. [46]	$PM_{2.5}$	
Liu et al. [47]	$PM_{2.5}$	attention + LSTM
Dairi et al. [48]	n/a	attention + VAE
Mengara et al. [49]	PM_{10}, $PM_{2.5}$	attention + Bi-LSTM
Liu et al. [50]	$PM_{2.5}$	attention + RNN
Elbaz et al. [51]	$PM_{2.5}$	attention + CNN/GRU
Abdullah et al. [52]	PM_{10}	MLP, RBF
Gradivsar et al. [53]	O_3	MLP
Shams et al. [54]	SO_2	MLP
Kurnaz et al. [55]	PM_{10}, SO_2	RNN
Chang et al. [56]	$PM_{2.5}$	LSTM
Song et al. [57]	AQI	RBN
Guo et al. [58]	API	ANN, WANN
Banasiewicz et al. [59]	NO_x	FFNN
Fernandez et al. [60]	NO_x	FAA-NN
Liu et al. [61]	NO_2	DWT + LSTM; GRU; Bi-LSTM
Mao et al. [62]	CO, NO_2, O_3, PM_{10}, $PM_{2.5}$, SO_2	GCN, LSTM
Abirami et al. [63]	CO, NO_2, NH_3, O_3, Pb, PM_{10}, $PM_{2.5}$, SO_2	hierarchical DL
Hickman et al. [64]	O_3	transformer
Zhang & Zhang [65]	$PM_{2.5}$	sparse attention transformer network
Xu et al. [66]	AQI	HGN
Iskandaryan et al. [67]	NO_2	AT-GCN
Santos et al. [68]	O_3	spatiotemporal graph network
Chen et al. [69]	AQI	GAGNN
Terroso-Saenz et al. [70]	C_6H_6, CO, NO, NO_2, NO_x, O_3, PM_{10}, $PM_{2.5}$, SO_2	GNN
Han et al. [71]	AQI, CO, NO_2, O_3, PM_{10}, $PM_{2.5}$, SO_2	SSH-GNN
Dua et al. [72]	NO_2, PM_{10}, $PM_{2.5}$	sequential modeling

Discussion

The presented studies confirm the increasing popularity and effectiveness of deep learning methods in forecasting air pollutant concentrations. Artificial neural networks, including multilayer perceptrons, remain a frequent choice due to their flexibility and relatively low computational cost, as shown in [30,31,54]. Several works demonstrate that the performance of such models, measured using statistical indicators such as R^2, RMSE, MAE, or MAPE, is often satisfactory and applicable in real-world environmental monitoring scenarios. Advanced architectures such as CNN-LSTM [42–44] and Bi-LSTM networks [41] are employed to capture both spatial and temporal dependencies in the data, which is particularly important when modeling particulate matter dynamics. Studies using LSTM-based approaches emphasize their strength in modeling long-term temporal dependencies and addressing data sequences with irregular patterns [56]. Some works introduce hybrid or optimized architectures, such as the use of genetic algorithms to enhance ANN performance by selecting the most relevant input variables [31]. The use of CNNs [40] and RNNs [55] further diversifies the methodological landscape, indicating that no single architecture dominates the field, but rather that the choice of method is often tailored to the specific pollutants, data availability, and local conditions.

The most frequently predicted pollutants include $PM_{2.5}$, PM_{10}, and SO_2, reflecting their importance in urban air quality assessment. Several authors emphasize the importance of meteorological and temporal features, such as wind speed, humidity, season, and day lag [54], indicating that the integration of environmental context is crucial for accurate forecasting. The wide range of architectures and pollutants studied demonstrates the adaptability of deep learning approaches, though it also highlights the need for standardized evaluation protocols and benchmark datasets to ensure comparability. Despite the progress, many models remain location-specific and dependent on the quality and resolution of available data.

Importantly, the potential applications of these models extend beyond academic research. High accuracy forecasts of pollutant concentrations can support the development of early warning systems, inform public health advisories, and guide short-term regulatory or traffic-control interventions in urban areas. In the longer term, insights derived from deep learning models may assist policymakers in evaluating the impact of environmental strategies and in optimizing resource allocation for air quality improvement.

A comparative analysis of deep learning methods for air quality forecasting reveals that the suitability of each architecture depends strongly on the characteristics of the available data and the forecasting horizon. CNN-based models are computationally efficient and particularly effective when high-resolution spatial features dominate, for example, in urban-scale pollution mapping. LSTM and Bi-LSTM networks, by contrast, excel at modeling sequential dependencies and capturing long-term temporal patterns, making them useful for predicting daily or seasonal pollutant variations. Transformer-based approaches are emerging as a promising alternative, especially for handling very long input sequences and multimodal datasets, though their practical deployment is still limited by high computational requirements and the need for large annotated datasets. Wavelet-enhanced networks occupy a more specialized niche, but their ability to denoise input signals and capture both time- and frequency-domain information makes them valuable when dealing with irregular or noisy sensor data.

From a practical perspective, CNNs are often applied in short-term urban forecasting tasks where spatial variability is crucial, while LSTMs and hybrid CNN–LSTM models are preferred for continuous monitoring and early warning systems that rely on temporal dynamics. Transformers may find broader adoption in large-scale environmental monitoring platforms that integrate ground sensors, satellite data, and meteorological inputs. Wavelet-based models, though less common, can support scenarios where the reliability of low-cost sensors is limited, as they enhance signal robustness under adverse conditions. Overall, no single architecture dominates; rather, their strengths and weaknesses are complementary. This diversity suggests that future research should emphasize the design of hybrid and ensemble frameworks that combine the efficiency of CNNs, the temporal learning capacity of LSTMs, the scalability of transformers, and the denoising capabilities of wavelets to achieve more accurate and generalizable forecasts across diverse environmental contexts.

Nevertheless, realizing the full potential of these approaches requires addressing several unresolved challenges. One major limitation is the interpretability of deep learning models, which often function as "black boxes". This can hinder trust in the results, particularly in policy making contexts where transparency is important. Recent work increasingly emphasizes the need for explainable artificial intelligence (AI) methods to address this issue. Additionally, access to high-quality, granular environmental data is not uniform globally. In many regions, deep learning approaches may be hindered by limited data availability or data privacy regulations, raising ethical concerns regarding equitable access to the benefits of AI-driven air quality forecasting.

4. Machine Learning Models

In this section, we present methods for forecasting air pollution concentrations based on machine learning. These techniques, including support vector machines (SVM), random forests (RF), linear regression (LR), ensemble learning (EL), or XGBoost, allow for effective modeling of the pollution levels. We will discuss the selected methods, their properties, and workflow.

In the study [73], a system for forecasting NO, NO_2, O_3, PM_{10} and $PM_{2.5}$ concentrations for 24 h ahead was developed. The methodology includes characterization, forecasting, and evaluation of the system using meteorological data and information on emission sources. A comparative analysis was conducted between artificial neural networks and the support vector model improved by particle swarm optimization, where SVR-PSO showed the highest forecasting performance. The system operates independently of the season, which confirms its stability and usefulness in air quality management.

In [74], an advanced warning system consisting of two modules is proposed: air quality forecasting and evaluation. The forecasting module uses a model combining data preprocessing and numerical optimization methods, as well as interval forecasting to take into account uncertainty. The evaluation module based on the cloud model enables the transformation of qualitative data into quantitative air quality assessments. The simulations conducted on data from the Chinese city of Dalian confirmed the high efficacy and wide applicability of the proposed system.

Air quality forecasting using machine learning to predict pollutant concentrations in hourly intervals is proposed in [75]. A multi-task learning approach is considered, taking into account meteorological data from previous days. A regularization method is introduced, enforcing similar forecast values in subsequent hours, which improves the model accuracy. The experiments show that the proposed techniques outperform standard regression models and existing regularization methods.

A study presented in [76] explored an ensemble learning method for predicting $PM_{2.5}$ concentrations. This approach utilized climate data and air quality metrics from adjacent cities. The predictive performance was evaluated using datasets containing $PM_{2.5}$ measurements from Beijing, Wuhan, and Shenzhen.

In [77], a machine learning model based on gradient boosting to forecast $PM_{2.5}$ concentrations in Taiwan is introduced. The study included a comparative analysis with data from Taipei and London, revealing similar predictive outcomes for both cities, likely due to their comparable topography. The model's performance was assessed using the coefficient of determination (R^2) and RMSE.

The research in [78] developed three machine learning models: multiple linear regression, random forest regression, and the general additive model for accurate forecasting of particulate matter concentrations. The results demonstrated high accuracy across all models, with the GAM model achieving the best predictive performance. Additionally, the study outlines a detailed methodology for predicting PM_{10} variability, offering a framework that can be adapted for different regions.

The study [79] uses the Prophet forecasting model to predict short-term and long-term air pollution in Seoul, including pollutants $PM_{2.5}$, PM_{10}, O_3, NO_2, SO_2, and CO. It outperforms similar models with significantly lower prediction errors, especially for $PM_{2.5}$ and PM_{10}. PFM's ability to predict air quality for up to a year shows promise for regions lacking advanced meteorological infrastructure. The study also emphasizes the potential for PFM to guide air quality management in Seoul.

The study [80] addresses the challenge of monitoring benzene (C_6H_6) in urban areas, where sensor deployment is expensive and limited. It proposes an ensemble machine learning approach to estimate C_6H_6 levels by utilizing the relationships between gases. Extensive experiments show that this technique significantly enhances the performance of existing machine learning methods for air pollution monitoring. The approach offers an efficient solution for benzene detection, potentially improving urban air quality monitoring without relying on costly sensor deployments.

The work [81] focuses on improving air quality and reducing human exposure to pollutants, particularly in developed cities in China, where ground-based measurement networks are too dispersed to capture spatial variability in rural and mixed-use areas. The authors utilize high-resolution satellite data, along with meteorological and chemical information from the WRF and CMAQ models, to estimate the vertical columnar NO_2 density and the distribution of near-surface NO_2 concentrations. They employ data assimilation methods and probabilistic kernel function models, with Gaussian regression providing the best results. They also analyze atmospheric CO_2 concentrations using data assimilation and inverse modeling, identifying high-emission areas and interannual variability in fossil fuel emissions in East Asia. The results allow for better monitoring and prediction of long-term changes in NO_2 and CO_2 concentrations and provide a baseline for implementing local pollution control measures and environmental policies.

In [82], a method for predicting the AQI is considered. The study includes data from the city of Visakhapatnam (India), covering 12 pollutants and 10 meteorological parameters. Several ML algorithms were tested, with CatBoost performing best, while AdaBoost obtained a lower R^2 value. The results confirm that machine learning techniques, especially CatBoost, enable highly accurate prediction of urban air quality and can be successfully applied globally.

The study [83] introduces a hybrid PSO–SVM model for short-term air pollutant concentration forecasting. The approach combines influence factor analysis, clustering of input data, and parameter optimisation to enhance predictive accuracy. A case study in Beijing demonstrates that the proposed method outperforms baseline and alternative models, including genetic algorithm-optimised SVM. The results highlight the potential of the hybrid framework as an effective tool for air quality management and early warning systems.

In the paper [84], an air quality forecasting system for Macau using five machine learning algorithms: ANN, RF, XGBoost, SVM, and MLR, to predict $PM_{2.5}$, PM_{10}, and CO concentrations over 24- and 48 h time horizons is developed. Measurement data were collected from 2016 to 2021, with 2020 to 2021 used for testing and the previous four years for model training. All models achieved good results for 24 h forecasts and acceptable accuracy for 48 h forecasts, although performance could be improved by feature selection using the SHAP test. The RF and SVM models achieved the best results for both $PM_{2.5}$, PM_{10}, and CO over 24- and 48 h forecasts.

The work [85] developed a machine learning model to forecast daily $PM_{2.5}$ concentrations in Shanghai by combining data from the operational numerical system WRF-Chem with ground-level measurements and meteorological conditions. The ML model significantly improved forecast accuracy compared to WRF-Chem alone, achieving 50–100% higher correlation coefficients and reducing the standard deviation. The results show that the use of machine learning can effectively complement chemistry-transport numerical models in air quality forecasting in China. This approach enables more precise pollutant predictions and supports better public health decisions.

A comprehensive literature review describing machine learning methods for modeling air quality may be found in [86]. The machine learning methods are summarized in Table 4.

Table 4. Machine learning methods for air pollutants forecasting.

Reference	Air Pollutants	Methods
Yu et al. [76] Kaur et al. [80]	$PM_{2.5}$ C_6H_6	EL EL
Murillo et al. [73]	NO, NO_2, O_3, PM_{10}, $PM_{2.5}$	PSO, SVM
Wang et al. [74]	CO, NO_2, O_3, PM_{10}, $PM_{2.5}$, SO_2	NOM
Zhu et al. [75]	O_3, $PM_{2.5}$	MTL
Lee et al. [77]	$PM_{2.5}$	GB
Eddine et al. [78]	PM	GAM, LR, RFR
Shen et al. [79]	CO, NO_2, O_3, PM_{10}, $PM_{2.5}$, SO_2	Prophet
Hugo [81]	CO_2, NO_2	n/a
Ravindiran et al. [82]	AQI	AdaBoost, CatBoost
Chen et al. [83]	CO, NO_2, O_3, PM_{10}, $PM_{2.5}$, SO_2	PSO-SVM
Lei et al. [84]	$PM_{2.5}$, PM_{10}, and CO	ANN, RF, XGBoost, SVM, and MLR
Ma et al. [85]	$PM_{2.5}$	XGBoost

Discussion

Machine learning methods have demonstrated considerable potential in the task of air pollution forecasting, particularly due to their ability to model complex, nonlinear relationships between meteorological variables, emission data, and pollutant concentrations. A variety of algorithms have been applied, including support vector machines (SVM), random forests (RF), gradient boosting methods such as XGBoost, multi-task learning, and time series approaches like Prophet. These methods are often combined with data preprocessing and optimization techniques, such as particle swarm optimization or regularization strategies, to improve forecasting accuracy. One of the notable strengths of machine learning models is their adaptability to different input features, such as meteorological data, historical pollutant levels, or even data from nearby regions, as seen in [76]. Moreover, architectures combining prediction with evaluation modules [74] or integrating interval forecasting [75] demonstrate that ML models can go beyond point estimation and support more robust decision making.

Each algorithm presents distinct advantages and limitations. SVM is valued for its robust generalization and interpretability, but can be computationally intensive for large datasets. RF provides high predictive accuracy and relatively fast training while offering feature importance insights, making it easier to interpret. XGBoost delivers state-of-the-art performance on structured data and handles complex nonlinear interactions effectively, although its hyperparameter tuning is more involved. Prophet offers ease of implementation and transparency in time series modeling, yet it may struggle with abrupt changes in pollutant concentrations. These differences illustrate that no single algorithm dominates the field, and the choice often depends on data availability, pollutant type, spatial resolution, and computational resources. Combining multiple ML methods or using ensemble strategies can further enhance forecast reliability and robustness.

The selection of specific machine learning models is often influenced by the characteristics of the target pollutant and the operational context. Particulate matter ($PM_{2.5}$, PM_{10}), which exhibits strong temporal and spatial correlations, benefits from models capable of capturing both sequential patterns and feature interactions, such as LSTM-enhanced or CNN–LSTM hybrid architectures. Gaseous pollutants like CO or NO_2, which show more localized and short-term variability driven by traffic or industrial emissions, are well suited to ensemble methods such as random forest or gradient boosting approaches like XGBoost, which effectively handle heterogeneous features and nonlinear dependencies. Time series–oriented models, including Prophet, are preferred for capturing seasonal patterns or when only coarse historical data is available. Moreover, hybrid or ensemble strategies are commonly adopted in complex urban environments to utilize the complementary strengths of multiple models, enhancing robustness under varying meteorological conditions, sensor coverage, or missing data. This context-specific selection underlines the importance of aligning algorithm choice with pollutant type, data characteristics, and operational requirements to achieve accurate and actionable air quality forecasts.

However, several challenges persist. First, model interpretability remains limited for many complex ML approaches, especially ensemble and black box models, which can hinder their acceptance by environmental agencies or policymakers. Second, generalization across seasons, locations, or cities is often assumed rather than empirically validated. Although [73] claims seasonal independence, such robustness is rarely tested across diverse contexts. Furthermore, while some studies utilize uncertainty quantification or multi-task settings, these techniques are not yet widespread. A significant gap also lies in the comparability of results, as the influence of input feature selection, data quality, and spatial resolution on model performance is often underexplored.

In summary, machine learning methods offer a powerful toolkit for forecasting air pollutants, balancing predictive performance, interpretability, and computational efficiency. Nevertheless, widespread deployment requires more rigorous validation, better explainability, and standardized evaluation practices. Future research should focus on integrating ML models with domain knowledge, heterogeneous sensor networks, and real-time data streams to enhance generalization, resilience, and actionable utility in environmental decision-making.

5. Statistical Models

In this section, we present statistical methods used to forecast air pollutant concentrations. Classical approaches, such as regression analysis, autoregressive models (AR), autoregressive models with moving average (ARMA), and their extensions (ARIMA, SARIMA), allow for the analysis and prediction of air quality changes based on historical data. These methods are particularly useful in modeling long-term trends and seasonal patterns of pollutant concentration changes. Although they are often characterized by lower accuracy than modern machine learning and deep learning models, their interpretability and lower computational requirements mean that they are still widely used.

The study [87] focuses on the analysis of $PM_{2.5}$ concentration in Fuzhou (China) using the ARIMA model. The results show that $PM_{2.5}$ shows seasonal fluctuations, with higher values in winter and lower values in summer. Spearman correlation analysis shows that $PM_{2.5}$ is positively correlated with PM_{10}, SO_2, and NO_2, and is negatively correlated with meteorological parameters. The forecasted values of $PM_{2.5}$ based on historical data suggest a decrease compared to previous years, which may be the result of policy actions taken by Fuzhou authorities.

In [88], the PM_{10} concentrations in Taiyuan (China) were analyzed using a wavelet-based ARMA/ARIMA model. The model's performance was assessed using multiple error measures. The results showed minimal values for the error metrics, indicating the model's high accuracy.

The work [89] examines the air quality in Lahore (Pakistan), focusing on the growing pollution issues that conflict with the pursuit of a high quality of life. The study reveals that $PM_{2.5}$ and PM_{10} concentrations exceed the National Environmental Standards (NEQS). Correlation analysis establishes the relationship between particulate matter and other pollutants such as O_3, NO, and SO_2, indicating that road traffic plays a significant role in the increase in emissions. A projection of future $PM_{2.5}$ concentrations using the SARIMA model predicts a further rise in particulate matter levels next year. The paper also provides a detailed analysis of air pollutants and their sources, stressing the need for actions to improve air quality in the region.

The research [90] explores the forecasting of monthly PM_{10} concentrations in Erzurum (Turkey) to take measures that minimize the risk of air pollution. The model was developed using the historical data. The data were cleaned from seasonal trends and effects and then decomposed into three levels. ARIMA models were applied to each subseries for forecasting the approximation and detail series. The WT-ARIMA model outperformed the traditional ARIMA in terms of RMSE, R^2, MAE, and MAPE metrics. The results demonstrate the high accuracy of the model, making it an effective tool for early warning in areas with high air pollution levels.

In [91], a statistical model was developed to predict $PM_{2.5}$ and PM_{10} concentrations, capturing up to 57.0% of the variability in $PM_{2.5}$ and 35.0% in PM_{10}. Temperature, wind speed, and wind direction explain 94.0% of the variability in $PM_{2.5}$, while relative humidity is particularly significant for PM_{10}. The inclusion of lagged PM concentrations further enhanced the prediction accuracy by 4.0–16.0%. When tested with historical data, the model showed strong correlations with observed PM concentrations and was effective in forecasting levels at additional monitoring sites.

In the work [92], the use of widely accessible LoRa devices and low-cost PM sensors for monitoring PM concentrations is discussed. The study also proposes a short-term forecasting method (up to 2 h) for PM concentrations using ARIMA and VARMA models. Experimental results from data collected by 15 LoRa sensors show a 7.77% improvement in forecast accuracy and a 3.7% increase in the correlation coefficient compared to forecasts made using data from a single sensor.

The paper [93] investigates the application of quantile regression to forecast extreme NO_2 concentrations in Madrid. Unlike point forecasts, this method allows for predicting the full probability distribution, which allows for better modeling of extreme values. The developed models, using meteorological data, outperform traditional point approaches in terms of accuracy and reliability. The importance of input variables is also analyzed, showing differences between the median and higher quantiles. Furthermore, a method for calculating the probability of exceeding threshold NO_2 values is presented.

In [94], ARMA and ARIMA models were used to predict daily mean concentrations of O_3, CO, NO, and NO_2 in Delhi. Data were subjected to a variance stabilizing transformation, and optimal model parameters were selected based on information criteria and autocorrelation functions. Accuracy was assessed using MAPE, MAE, and RMSE. The results indicate the effectiveness of this method in short-term air quality forecasting.

The study [95] proposes a stepwise multiple linear regression model to forecast PM_{10} and NO_x concentrations in Athens and Helsinki, using pollutants and meteorological variables as predictors. Results showed that including Monin–Obukhov length and mixing height improved forecast accuracy, especially for Helsinki. MLR outperformed artificial neural networks in some cases, providing reliable predictions for daily averages and maximum hourly concentrations. The study concluded that MLR models are a useful tool for regulatory air quality forecasting despite limitations, such as the inability to capture rapid meteorological changes.

In the work [96], the ARIMA model to predict atmospheric environmental quality in Hunan Province (China) is proposed. The findings suggest a year-on-year improvement in air quality, with forecasts indicating gradual enhancement. The model proves reliable for predicting air pollutants ($PM_{2.5}$, PM_{10}, SO_2, CO), providing valuable insights for future air quality management in the region.

The statistical methods for air pollutants forecasting are summarized in Table 5.

Table 5. Statistical methods for air pollutants forecasting.

Reference	Air Pollutants	Methods
Zhang et al. [87]	$PM_{2.5}$	ARIMA
Gao et al. [96]	CO, PM_{10}, $PM_{2.5}$, SO_2	ARIMA
Zhang et al. [88]	PM_{10}	ARMA, ARIMA
Kumar et al. [94]	CO, NO, NO_2, O_3	ARMA, ARIMA
Yun et al. [92]	PM	ARIMA, VARMA
Bhatti et al. [89]	NO, O_3, PM_{10}, $PM_{2.5}$, SO_2	SARIMA
Aladaug et al. [90]	PM_{10}	WT-ARIMA
Afrin et al. [91]	PM_{10}, $PM_{2.5}$	statistics
Aznarte et al. [93]	NO_2	QR
Vlachogianni et al. [95]	NO_x, PM_{10}	MLR

Statistical models each come with specific assumptions, typical applications, and inherent limitations. ARIMA models assume stationary time series and linear relationships, making them suitable for short-term pollutant forecasts in urban areas, but they struggle with nonlinear interactions and multivariate data. SARIMA extends ARIMA by modeling seasonal patterns, which is beneficial for regions with strong seasonal effects, yet it requires careful parameter tuning and is less adaptable to high-dimensional datasets. Multiple linear regression (MLR) assumes linearity and independence of predictors, providing rapid estimation across multiple features, though it is sensitive to multicollinearity and less effective for capturing complex nonlinear dynamics. Considering these aspects highlights why statistical models remain relevant and useful, especially in low-data or high-interpretability scenarios, despite the increasing popularity of machine learning and deep learning approaches.

Discussion

Statistical models, particularly those based on autoregressive techniques, have long been applied in air quality forecasting due to their conceptual simplicity, transparency, and interpretability. Techniques such as ARIMA, SARIMA, and generalized additive models (GAM) allow for straightforward modeling of temporal trends and seasonal patterns, as demonstrated in studies like [87–89]. These models provide both accurate forecasts of pollutant concentrations and valuable insights into interrelationships among pollutants and environmental variables. Their interpretability is especially advantageous in policy-making contexts, where understanding the influence of meteorological factors, emission sources, and seasonal effects is crucial.

Despite the rise in machine learning and deep learning approaches, statistical models remain relevant in several scenarios. They require relatively low amounts of data and computational resources, making them suitable for regions with limited monitoring infrastructure, such as parts of Africa or Southeast Asia. In heavily polluted regions like China or India, these models are often used to establish baselines and capture dominant seasonal cycles (e.g., winter smog episodes), while in European or North American contexts, they are valued for their transparency, supporting regulatory compliance and communication with stakeholders. Moreover, they serve as robust benchmarks for evaluating more complex models and can be integrated into hybrid frameworks that combine statistical rigor with data-driven enhancements. Their main limitations include difficulties in capturing nonlinear interactions, limited adaptability to high-dimensional or multivariate datasets, and challenges in utilizing spatial information, which are the issues that are particularly pronounced in regions with complex land use patterns (e.g., urban-industrial corridors or coastal zones) where pollutant dispersion is highly variable.

In addition to the aforementioned considerations, statistical models play distinct roles depending on the forecasting horizon and application. For nowcasting tasks, they are valuable for short-term air quality alerts and rapid identification of pollution peaks, especially in urban areas with limited real-time monitoring. For forecasting and prediction, they help capture seasonal cycles, trends, and long-term pollutant dynamics, providing essential baselines against which machine learning or hybrid models can be benchmarked. In particular, their interpretability facilitates communication with policymakers and stakeholders regarding the expected impact of emission control measures.

The choice of numerical schemes and model structures is also critical. For instance, AR, ARMA, and ARIMA are effective for univariate time series with temporal autocorrelation, while SARIMA utilizes seasonal components. GAMs allow flexible nonlinear relationships between predictors and pollutants, and VARMA or wavelet-enhanced ARIMA address multivariate dependencies and nonstationarity. Moreover, downscaling techniques can enhance the spatial resolution of forecasts, allowing statistical models to use localized meteorological or land-use information, which is especially important in regions with complex terrain or heterogeneous emission sources. These methodological choices determine model performance, robustness, and applicability across different geographical and environmental contexts.

Nevertheless, statistical models can be extended and improved in various ways. Approaches such as wavelet-enhanced ARIMA [90] or multivariate structures like VARMA [92] have shown promise in addressing nonstationarity and multivariate dependencies, although these methods are still underutilized compared with machine learning techniques. Furthermore, their strong reliance on temporal autocorrelation means that careful preprocessing is often necessary, especially in heterogeneous urban environments or areas with strong meteorological variability, such as mountainous regions. Despite these challenges, the simplicity and transparency of statistical methods make them an indispensable component of the air quality forecasting toolbox, particularly in low-data or high-interpretability scenarios.

To sum up, statistical methods remain a valuable benchmark for evaluating the performance of more advanced models. They are especially appropriate in data-scarce environments or as part of hybrid systems combining statistical foundations with machine learning enhancements. Future research may benefit from integrating statistical rigor with the flexibility of modern data-driven approaches, thereby improving robustness and interpretability in air quality forecasting across diverse geographical contexts and environmental settings.

6. Hybrid Models

In this section, we present hybrid methods that combine different approaches to improve the performance of air pollutant concentration forecasting. A hybrid model in the context of air quality forecasting can be defined as a framework that combines two or more modeling paradigms (e.g., statistical, machine learning, physical, or optimization-based approaches) in order to utilize their complementary strengths while compensating for individual weaknesses [97]. For example, ARIMA models can be used to capture temporal trends, while neural networks or XGBoost algorithms can better handle nonlinearities in the data. Another popular approach is to combine classical regression methods with deep learning models, which allows for a more precise representation of the relationships between meteorological factors and air quality.

6.1. Deep Learning + Statistical Methods

The study [98] explores statistical regression and computational intelligence-based models for forecasting hourly NO_2 concentrations at the Taj Mahal in Agra, with an emphasis on public health-related air quality predictions. The research utilized historical air pollution data, focusing on months when pollution levels peaked. The forecasting models were developed using multiple linear regression and a principal component analysis combined with an artificial neural network method. The proposed model outperformed multiple linear regression, showing higher correlation coefficients and better statistical alignment, and proving to be a reliable tool for predicting air pollution levels at the Taj Mahal.

The paper [99] explores various machine learning techniques for predicting PM_{10} concentrations in Lublin, Poland. The methods assessed include linear regression, *k*-nearest neighbors regression, support vector machine, regression trees, Gaussian process regression, artificial neural network, and long short-term memory networks. Based on the coefficient of determination (R^2), the experimental analysis found that the artificial neural network provided the most accurate predictions among the methods tested. Further studies on air quality analysis in Poland are available in [100,101].

In [102], a hybrid approach for forecasting $PM_{2.5}$ concentrations is proposed. This method combines a gradient-boosted regression tree with convolutional neural networks and fuzzy *k*-nearest neighbor. The forecasting results were compared with several methods, including multiple linear regression, stacked long short-term memory, bi-directional gated recurrent unit, and gradient-boosted regression tree. Evaluation based on metrics such as root mean square error, mean absolute error, and mean absolute percentage error demonstrated that the proposed method outperforms the others in prediction accuracy.

The work [103] addresses urban air quality forecasting, particularly focusing on the complex temporal patterns that are challenging to predict, especially during extreme events. The authors propose a hybrid model that combines autoregressive integrated moving average (ARIMA) and artificial neural networks to enhance forecast accuracy. The model is applied in Temuco (Chile), where firewood burning is the primary source of PM_{10} pollution during the winter. Meteorological data and PM_{10} measurements are used in the model. Experimental results indicate that the hybrid model outperforms individual models such as ARIMA, artificial neural network, and multilinear regression, successfully forecasting 100.0% of alarm episodes and 80.0% of pre-alarm episodes. This approach demonstrates its potential for application in air quality forecasting across various cities and countries.

In [104], effective methods for forecasting PM_{10} and $PM_{2.5}$ are examined. The study compares various forecasting models, including neural networks (multilayer perceptron, radial basis function, and echo state networks), as well as statistical methods such as the autoregressive model and autoregressive moving average model. The best forecasting results were achieved by combining different types of neural networks.

In [105], NO_2 concentrations were predicted at 14 measuring stations in the UAE. Statistical and machine learning models were used: ARIMA, SARIMA, LSTM, and NAR-NN in open and closed-loop architectures. Accuracy was assessed using MAPE, and it was found that MAPE strongly correlates with the relative standard deviation of NO_2 concentrations.

The study [106] compares multilinear regression and multilayer perceptron in predicting NO_2 concentration in Tehran's air based on meteorological data, urban traffic, and green areas. The results show that the multilayer perceptron obtains higher forecasting accuracy. Sensitivity analysis indicates that green areas have a key impact on NO_2 reduction, even surpassing traffic density. A multilayer perceptron is proving to be an effective tool for modeling complex environmental interactions.

The paper [107] proposes a hybrid model combining an artificial neural network and a support vector machine algorithm to improve air quality forecasting accuracy. The method enhances traditional ANN and SVM predictions by using a Taylor expansion-based error correction approach. The models utilize historical air pollution data and meteorological variables to forecast PM_{10} and SO_2 concentrations in Taiyuan (China). Experimental results using a two-year dataset demonstrate that the hybrid models effectively reduce forecasting errors, offering a promising approach for improving air quality predictions in urban environments.

In [108], an evaluation of three forecasting methods: Elman neural networks, ARIMA models, and a hybrid approach combining both to predict SO_2 pollution episodes near a coal-fired power station is provided. The hybrid model outperformed the individual methods, achieving the best forecast accuracy. This predictive capability is crucial for power plants, allowing them to take preventive measures to reduce emissions and decrease environmental impact.

The research [109] discusses a long short-term memory network to predict NO_x emissions in Pohang (South Korea), using stochastic regression for missing data imputation. The model's performance is evaluated through the mean absolute scaled error, indicating its ability to outperform a simple naive prediction. The study focuses on optimizing parameters like time windows and learning rates, making it a valuable tool for companies to meet regulatory quotas and reduce costs.

In [110], an evaluation of machine learning models for air pollution forecasting, focusing on support vector regression, ARIMA, and long short-term memory, is considered. Using time series data for PM_{10} and $PM_{2.5}$, the models were tested for a predictive accuracy. Results indicate that SVR and ARIMA are the most effective for forecasting air pollutant concentrations.

In the work [111], an improved air quality index (AQI) prediction model utilizing historical and current meteorological datasets is presented. It evaluates various classifiers on India's Air Pollution Geocodes Dataset. The decision tree classifier achieved 99.7% accuracy, improving slightly with the random forest classifier. Forecasting was conducted using ARIMA, providing AQI projections for the next 45 days and monthly forecasts for a year. The study highlights how integrating multiple models enhances accuracy, with decision trees and ensemble methods yielding the best results.

6.2. Deep Learning + Optimization and Metaheuristics

The work [112] presents a multi-stage forecasting approach for $PM_{2.5}$ concentrations using a novel hybrid model that integrates various machine learning techniques. The proposed method combines wavelet packet decomposition, particle swarm optimization, back propagation neural network, and the AdaBoost algorithm. The results showed that the usage of PSO and AdaBoost significantly improves forecasting accuracy over other methods.

The study [113] explores the forecasting of air pollution, which can assist urban planning for maintaining a sustainable environment and reducing health risks. The study focuses on predicting $PM_{2.5}$ concentration at two air monitoring stations in Kuala Lumpur (Malaysia), using hybrid deep learning models. These models, based on 4 h data, utilize six types of pollutants, meteorological parameters, and data from nearby stations. Particle swarm optimization (PSO)- and sparrow search algorithm (SSA)-optimized LSTM models (PSO-LSTM, SSA-LSTM) were developed to identify the most significant combinations of input data. The selected configurations were then applied to empirical mode decomposition (EEMD) to create models: EEMD-PSO-LSTM and EEMD-SSA-LSTM. These models showed notable improvements in forecast accuracy. Among these, the EEMD-SSA-LSTM model outperformed the others, yielding better RMSE and MAE results at the chosen air monitoring stations. The analysis demonstrated that utilizing data from nearby $PM_{2.5}$ stations significantly enhances prediction accuracy, and SSA was found to be more effective than PSO in optimizing LSTM parameters.

In [114], an adaptive neuro-fuzzy inference system combined with a genetic algorithm to model NO_x emissions from a natural gas-fired combined cycle power plant is utilized. The model was trained using emission data, showing high predictive accuracy. The results suggest that the genetic algorithm significantly improves the model's performance, reducing error and enhancing predictive accuracy. The model demonstrated potential for accurate NO_x emissions forecasting, with minimum error values across multiple criteria.

The work [115] introduces a hybrid intelligent model combining long short-term memory and the multi-verse optimization algorithm (MVO) to predict NO_2 and SO_2 emissions from a Combined Cycle Power Plant. The LSTM model acts as a forecasting engine, while MVO optimizes its parameters to reduce prediction errors. Tested on real-world data from a power plant in Kerman (Iran), the model outperformed existing benchmark models, demonstrating higher accuracy under different input configurations, including meteorological factors and lagged pollutant values.

In [116], hybrid models were proposed for forecasting daily concentrations of various particulate matter fractions and the Air Quality Index (AQI) in Romania. The methods employed included input variable selection, machine learning, and regression models. Evaluation using measures such as R^2 and RMSE demonstrated satisfactory results in forecasting PM concentrations.

6.3. Ensemble Methods

The work [117] compares the performance of a multilayer perceptron and support vector regression in forecasting NO and NO_2 concentrations based on meteorological data and historical measurements in Szeged (Hungary). Principal component analysis was used to reduce the dimensionality of the data, and the grid search method was used to optimize the model parameters. The results indicate that nonlinear methods outperform linear methods, especially in the case of NO. For NO_2, the improvement is noticeable but less pronounced.

In [118], a two-step methodology to forecast tropospheric ozone (O_3) levels in the Sines region of Portugal is developed. It combined meteorological, air quality, and industrial emissions data to predict O_3 concentrations up to 24 h in advance. The approach used classification and regression trees to identify key predictors and multilayer perceptron models for forecasting. The results showed high prediction accuracy, offering the potential for real-time health and environmental advisories.

The study in [119] presents a hybrid forecasting model that integrates a neural network with the NARX model and uses automatic feature selection. This approach enhances the accuracy of $PM_{2.5}$ predictions, particularly for peak concentrations, and provides a clearer understanding of the influence of weather and seasonal conditions. The experimental results demonstrate strong model performance, with high values of correlation coefficients for one-hour forecasts. The proposed method offers more reliable air quality forecasting and improves the interpretation of factors contributing to pollution.

The paper [120] investigates hybrid approaches for short-term forecasting of $PM_{2.5}$ and PM_{10} concentrations in major Polish cities. The study utilizes stepwise regression, tree-based algorithms (random forests and XGBoost), and neural networks. The experimental results validate the high accuracy of these forecasting techniques.

In the study [121], a hybrid model combining neural networks is introduced for predicting $PM_{2.5}$ concentrations. The model integrates graph convolutional networks with long short-term memory networks to forecast the spatiotemporal changes in $PM_{2.5}$ concentrations. Compared with other leading methods, the discussed model demonstrated higher forecasting accuracy in terms of various machine learning metrics, including the correlation coefficient R^2.

The work [122] introduces a two-step hybrid model for forecasting NO_2 and SO_2 concentrations in four cities in Central China. The model decomposes pollutant data using complete ensemble empirical mode decomposition (CEEMD), then applies support vector regression (SVR) combined with Cuckoo Search (CS) and Grey Wolf Optimizer (GWO) to predict high- and low-frequency components. The final forecast is obtained by summing both predictions. Evaluation results using MAE, MAPE, and RMSE confirm that the CEEMD-CS-GWO-SVR model outperforms other forecasting approaches in accuracy.

6.4. Boosting Models

In [123], the particle swarm optimization and artificial bee colony techniques to forecast CO_2 emissions in Turkey, using socioeconomic indicators such as energy consumption, population, and the number of vehicles, are described. Models are developed in linear, exponential, and quadratic forms, utilizing CO_2 historical emissions data. The sum of squared error is used as a fitness function. In [124], a sensor array combined with a multilayer perceptron and long short-term memory model to enhance gas concentration estimation in outdoor environments, addressing issues like temperature, humidity, and cross-sensitivity interference, is used. The experimental results demonstrate improved accuracy in predicting CO, NO_x, NO_2, and C_6H_6 concentrations, in terms of R^2 measure, showcasing the efficacy of the MLP-LSTM model in gas sensing applications.

The study [125] proposes a hybrid air pollution forecasting model integrating stacking-based ensemble learning and deep learning frameworks. The approach combines multiple machine learning models, including gradient-boosted tree regression, support vector machine-based regression, and long short-term memory to enhance predictive accuracy for air quality. Experimental results demonstrate the hybrid model's advantage over traditional models, highlighting its efficacy in improving air pollution forecasting performance.

6.5. Other Hybrid Methods

The research [126] discusses a new hybrid forecasting system for $PM_{2.5}$ concentrations. The system begins with an efficient data analysis method to extract the main features from the $PM_{2.5}$ dataset while minimizing noise influence. The Harris hawk optimization algorithm is then used to tune the extreme learning machine model, which achieves high prediction accuracy. The health impacts and economic costs of pollution are estimated based on the predicted $PM_{2.5}$ levels. Experiments conducted with real data from Beijing, Tianjin, and Shijiazhuang show that the proposed system can support environmental management, emission reduction efforts, and health prevention strategies. It is also applicable to various fields, such as $PM_{2.5}$ health research.

In the study [127], a support vector machine kernel designed for predicting the atmospheric pollutant factor PM_{10} in time series problems is presented. The proposed kernel uses a transformed particle swarm-optimized artificial neural network weight vector within a Bayesian-optimized SVM kernel. Experimental results demonstrate that such a new kernel significantly improves forecasting accuracy compared to conventional ANN and SVM models. The findings suggest that the proposed SVM kernel offers an enhanced forecasting technique for air pollution prediction. The machine learning algorithms were also analyzed in [128].

In [129], a time series modeling to predict the concentrations of major pollutants using the adaptive neuro-fuzzy inference system, which better accounts for nonlinear relationships than traditional models, is proposed. The concentrations of CO, SO_2, O_3, and NO_2 data were used for evaluation. The results show that the coefficients of determination for the proposed model are higher than those obtained for the semi-experimental model. The improved forecasting accuracy can support the development of effective environmental and public health policies, contributing to sustainable development.

The study [130] addresses NO_x measurement delays in coal-fired power plants, which lead to issues in ammonia injection and compliance with environmental standards. A hybrid boosting model is proposed to predict NO_x levels accurately and compensate for measurement delays, improving ammonia injection for denitrification. The model, tested on actual plant data, outperforms ARIMA in short-term forecasts, with reduced errors. This model is suitable for real-time on-site implementation as a soft sensor for power plants.

In [131], a combination of the random forest algorithm with NASA's GEOS-CF product to forecast O_3 and NO_2 concentrations in southeastern China is discussed. The model provides continuous, spatiotemporal forecasts for up to five days, significantly improving upon the initial GEOS-CF model by reducing errors. It offers a more accurate and comprehensive approach to monitoring air quality, which is crucial for decreasing the harmful effects of these pollutants.

The research [132] proposes a hybrid prediction method combining empirical mode decomposition (EMD), a transformer, and a bidirectional LSTM (Bi-LSTM) network, designed for ultrashort-term forecasts of nonlinear time series. The AQI data is first decomposed into intrinsic mode functions (IMFs), then each IMF is predicted separately using the improved transformer-Bi-LSTM approach, with linear prediction applied to simpler trends. Finally, the predicted IMFs are integrated using Bi-LSTM to produce AQI forecasts, achieving low RMSE, MAE, and MAPE values over a 5 h window. Validation on datasets from multiple other cities demonstrates that the model is both highly accurate and broadly applicable for real-time air quality prediction.

A review of hybrid methods for air quality forecasting is also provided in [133–138]. The hybrid methods for air pollutants forecasting are summarized in Table 6.

Table 6. Hybrid methods for air pollutants forecasting.

Reference	Air Pollutants	Methods
Mishra et al. [98]	NO_2	MLR, PCA, PCA + ANN
Kujawska et al. [99]	PM_{10}	*k*-NNR, ANN, GPR, LR, LSTM, RT, SVM
Usha et al. [102]	$PM_{2.5}$	GBTR + CNN + fuzzy *k*-NN
Diaz et al. [103]	PM_{10}	ARIMA + ANN
Neto et al. [104]	PM_{10}, $PM_{2.5}$	ESN + AR, ARMA; MLP, RBF
Al et al. [105]	NO_2	ARIMA, LSTM, NAR-NN, SARIMA
Shams et al. [106]	NO_2	MLP, MLR
Wang et al. [107]	PM_{10}, SO_2	ANN + SVM
Sanchez et al. [108]	SO_2	ARIMA, ENN
Lee et al. [109]	NO_x	LSTM + SR
Dobrea et al. [110]	PM_{10}, $PM_{2.5}$	ARIMA, LSTM, SVR
Sharma et al. [111]	AQI	ARIMA, DT, RF
Liu et al. [112]	$PM_{2.5}$	AdaBoost, ANN, PSO, WPD
Ahmed et al. [113]	$PM_{2.5}$	PSO- and SSA-optimized LSTM
Dirik et al. [114]	NO_x	ANFIS + GA
Heydari et al. [115]	NO_2, SO_2	LSTM + MVO
Udristioiu et al. [116]	PM	input variable selection, RM
Juhos et al. [117]	NO, NO_2	MLP, SVR
Durao et al. [118]	O_3	MLP, RT
Gu et al. [119]	$PM_{2.5}$	ANN + NARX
Czernecki et al. [120]	PM_{10}, $PM_{2.5}$	ANN, RM, RF, XGBoost
Qi et al. [121]	$PM_{2.5}$	GCN + LSTM
Zhu et al. [122]	NO_2, SO_2	complete ensemble empirical mode decomposition, SVR + Cuckoo Search and Grey Wolf Optimizer
Ozceylan et al. [123]	CO_2	ABC, PSO
Chang et al. [125]	PM_{10}, $PM_{2.5}$	GBTR, LSTM, SVM
Du et al. [126]	$PM_{2.5}$	extreme learning machine model
Kouziokas et al. [127]	PM_{10}	new kernel for SVM
Zeinalnezhad et al. [129]	CO, NO_2, O_3, SO_2	ANFIS
Lyu et al. [130]	NO_x	boosting model
Li et al. [131]	NO_2, O_3	RF + GEOS-CF
Dong et al. [132]	AQI	EMD + transformer-Bi-LSTM
Ng et al. [124]	C_6H_6, CO, NO_x, NO_2	sensor array + MLP, LSTM
Harishkumar et al. [128]	PM	DT, GBR, MLP, RF

Discussion

Hybrid methods used to forecast air pollutant concentrations offer significant advantages over traditional forecasting techniques. They combine the power of different algorithms, such as neural networks, regression, principal component analysis, and optimization techniques, to improve the accuracy and interpretability of the results. Hybrid models, such as ANN with NARX in [119] or PCA with ANN in [98], show higher correlation coefficients and better forecast accuracy compared to single methods. These models are particularly effective in predicting pollution during peak periods, where the quality of forecasts is crucial for protecting public health.

The diversity of hybrid approaches in air pollution forecasting can be organized into four main categories. Deep learning + statistical hybrids (e.g., ARIMA + LSTM) utilize the interpretability of classical models while benefiting from the nonlinear learning capacity of DL architectures. Deep learning + optimization and metaheuristics approaches (e.g., PSO- or MVO-optimized LSTM) focus on parameter tuning and convergence, typically yielding performance gains in predictive accuracy and training efficiency. Ensemble methods combine multiple learners, such as RF, SVM, or ANN, to improve robustness and reduce variance, making them suitable for heterogeneous or noisy datasets. Finally, boosting models like XGBoost or gradient boosting trees have gained popularity due to their strong accuracy in capturing complex nonlinearities, albeit with higher training costs and reduced interpretability. This categorization both clarifies the methodological landscape and also emphasizes the trade-offs researchers face when selecting hybrid approaches.

Hybridization improves forecasting performance primarily because it allows different methods to compensate for each other's weaknesses. Statistical models provide interpretability and capture seasonality or long-term trends, while machine and deep learning models excel at modeling nonlinearities and high-dimensional feature interactions. Optimization and metaheuristic techniques further refine model parameters, leading to faster convergence and reduced risk of local minima. Ensemble and boosting strategies increase robustness by combining diverse learners, which helps to prevent overfitting and enhances generalization. Such synergies are particularly valuable in scenarios where pollutant dynamics are influenced by multiple interacting factors, such as rapidly changing meteorological conditions, mixed emission sources, or heterogeneous spatial settings. In these contexts, hybrid approaches often outperform single models by delivering more accurate, resilient, and context-sensitive forecasts.

Despite these advantages, there are several challenges associated with the implementation of hybrid methods. First of all, the development of such models requires access to large, high-quality data sets, which can be problematic, especially in countries with limited access to pollution data. In addition, the selection of appropriate algorithms and optimization methods requires a thorough analysis of the specifics of a given region and the type of pollutants. It should be noted that each component of the hybrid model may introduce additional complexities and risks of overfitting, especially when data are limited. Another challenge is the interpretation of results, especially in the context of complex, nonlinear relationships between meteorological variables and pollution levels. Although methods such as ANFIS [129] allow for better capturing of these relationships, there is still a need for further development of tools for visualization and explanation of forecasts that will allow their practical application in air quality management. Additionally, many of the presented models rely on advanced computational techniques that require large computational resources, which may be a barrier to their wider implementation in real-time operational systems. This requires the use of computational technologies that can meet the requirements for both speed and accuracy of forecasts.

Moreover, in the context of air pollution modeling, it is necessary to take into account local and seasonal variability, which is a significant challenge, especially in cities with dynamic atmospheric conditions. Although hybrid models offer advanced approaches, much remains to be done to develop forecasting systems that can take into account local conditions in a scalable manner and are robust to changes in input data.

In general, hybrid models that integrate statistical approaches with machine learning tend to perform better than either approach alone, as they can capture both linear dependencies and complex non-linear patterns. For regions with highly dynamic or heterogeneous environments (e.g., mixed urban–industrial areas or complex topographies), hybrid methods that combine physical or chemical transport models with data-driven learning are often more robust. However, the relative advantage of each type of hybrid depends strongly on data availability and the specific environmental context, which suggests that no single hybrid architecture can be universally recommended.

In summary, hybrid methods are a promising alternative to traditional air quality forecasting models, but their effectiveness largely depends on the quality of data, the appropriate selection of methods and algorithms, and the ability to interpret results in the context of changing atmospheric conditions.

7. Other Methods

The study in [139] examines hourly $PM_{2.5}$ concentration forecasting in southwestern Poland using the uEMEP and EMEP4PL models. Forecast accuracy is assessed by comparing the results with data from measurement networks. EMEP4PL, an adaptation of the EMEP MSC-W chemistry transport model for Poland, operates with two nested domains: a European grid (12 km $\times$ 12 km) and a more refined Polish grid (4 km $\times$ 4 km). The uEMEP model, employing Gaussian modeling principles, further downscales EMEP4PL results using the local fraction approach. Findings reveal that uEMEP forecasts exhibit lower bias and a higher index of agreement than EMEP. Additionally, [140] discusses the use of a random forest algorithm to enhance daily $PM_{2.5}$ concentration estimates produced by the EMEP4PL model in Poland.

The work [141] explores short-term forecasting of PM_{10} concentrations in Krakow using sodar Doppler technology, with a focus on the influence of meteorological conditions on air pollution levels. The proposed approach utilizes the spectrum representing the amplitudes of signals reflected back to the sodar receiver from a single-frequency sound transmission, whose properties vary based on atmospheric conditions. Data mining techniques were applied to predict episodes of high PM_{10} concentrations. In [142], particulate matter concentrations in selected Polish cities, including Wrocław and Poznań, from 2014 to 2016 were analyzed. The analysis revealed that the highest concentrations of $PM_{1.0}$, $PM_{2.5}$, and PM_{10} occurred in 2016 in both cities. The study [143] proposes a method for predicting the number of days with atmospheric visibility in Warsaw, Poland, using meteorological and air quality data. The approach employs the Iman–Conover method to simulate a time series of meteorological and air quality parameters. Statistical validation confirmed that the model accurately predicts the number of days within a month when visibility falls within specific ranges.

In [144], a method for predicting a comprehensive air quality index in Xingtai (China), utilizing the speed of socio-economic development in the forecasting model, is presented. The findings demonstrate that the grey multivariable model employed in the study achieves high predictive accuracy.

The study [145] presents an adaptive model that combines artificial neural networks with wavelet analysis for forecasting daily $PM_{2.5}$ concentrations using remote sensing data and ground observations. The model's performance was evaluated using the Pearson correlation coefficient, mean absolute percentage error, root mean square error, and mean absolute error. The results demonstrate that the wavelet model outperforms the ANN, showing higher forecast accuracy, better stability, and smaller errors during both the training and validation periods. These findings confirm the efficacy of the WANN model for predicting $PM_{2.5}$ concentrations one day in advance. The daily forecasting is also discussed in [146].

In [147], the PM_{10} concentrations during the winter months (December–January) from 2009 to 2015 in Poland were analyzed. The findings indicate that low wind speeds contribute to higher levels of air pollution. The study used data from three monitoring stations in the region, allowing for the identification of smog episodes (referred to as black smog) and the verification of measurements at a station located near a large incineration plant. The results confirm that wind direction plays a significant role in the movement of pollutants and their local deposition.

The paper [148] focuses on forecasting particulate matter concentrations to aid regulatory planning, emission reduction, and the implementation of early warning systems. The authors introduce the CS-EEMD-BPANN model, which integrates gray correlation analysis (GCA), ensemble empirical mode decomposition (EEMD), cuckoo search algorithm (CS), and backpropagation neural networks (BPANN). A key innovation is the use of GCA to identify potential predictors of particulate matter from other air pollutants (such as CO, NO_2, O_3, SO_2) and meteorological variables (including wind speed and direction, temperature, humidity, and pressure). The model was tested in four Chinese cities with diverse climatic conditions, terrain, and emission sources—Beijing, Shanghai, Guangzhou, and Lanzhou. The analysis found that CO, NO_2, and SO_2 are strongly correlated with PM, and their inclusion significantly enhances the model's performance, confirming the effectiveness of the proposed approach. The analysis of CO_2 emissions is considered in [149].

The paper [150] describes an upgrade to the NAQFC system aimed at improving $PM_{2.5}$ forecasts. The system now includes real-time emissions from wildfires and windblown dust, reductions in dust emissions from snow- or ice-covered areas, and a shortened life cycle of organic nitrates in the gas chemistry. Testing and evaluations conducted over multiple seasons, compared with the EPA AIRNow monitoring network, showed significant improvements in regional forecasts. For instance, including windblown dust improved forecasts in the Western States, and reducing winter dust emissions resulted in a 52 percent reduction in forecast bias in the north-central United States.

In the study [151], a synoptic meteorological classification was performed to analyze its correlation with particulate matter concentrations at five sampling points in Valladolid (Spain), using a Q-mode clustering technique. The analysis considered 11 meteorological variables and the 500-hPa contour, identifying five synoptic meteorological types, which were further divided into nine synoptic classes. The algorithm was also applied to particulate matter concentrations, resulting in seven pollution levels for each sampling point. The comparison between synoptic types and pollution levels revealed a strong correlation between meteorological conditions and particulate matter concentrations at the 500-hPa contour.

In [152], a dynamic data-driven model for estimating NO_x and CO emissions from coal-fired utility boilers, focusing on forecasting emissions, is presented. The model demonstrates higher accuracy compared to static models and can be used in dynamic optimization algorithms to minimize emissions. The results indicate that applying both dynamic and steady-state optimization leads to lower emissions compared to historical plant data.

The study [153] evaluates the impact of assimilating surface ozone and fine aerosol measurements on forecast accuracy. Using the Weather Research and Forecasting–Chemistry model along with the Grid-point Statistical Interpolation tool, an assimilation experiment was conducted over northeastern North America, utilizing hourly data from the AIRNow network. Background error covariance was derived from July 2004 forecasts, and model performance was assessed using forecasts from August and September 2006, both with and without assimilation. The results show that assimilation improves ozone and fine aerosol forecasts, leading to enhanced verification scores for at least 24 h. However, this improvement is partially counterbalanced by rapid model error growth during the initial hours of forecasting.

In [154], forecasting CO_2 emissions in China's cement industry using a novel grey prediction model, accounting for uncertainty, is considered. The proposed model, V-GM, demonstrated 97.29% accuracy in simulating historical emissions data. It offers a more flexible approach compared to traditional methods, with improved forecasting capabilities. The model also includes an uncertainty analysis, making it useful for decision makers in achieving reliable, cost-effective predictions for emission decrease and sustainable development.

The work [155] examines the hygroscopic properties of particulate matter (PM), which influence light scattering and absorption. The study investigates the behavior of coarse PM (CPM) and fine PM ($PM_{2.5}$) under varying weather conditions and their effect on visibility. A censored regression model was developed to analyze the relationship between PM concentrations and meteorological factors, which facilitated the calculation of the optical hygroscopic growth factor and hygroscopic mass growth. These metrics were applied to $PM_{2.5}$ field measurements using low-cost sensors in two distinct regions. The findings show that CPM and $PM_{2.5}$ concentrations negatively affect visibility, with relative humidity (RH) significantly modulating these impacts. The modeled hygroscopic growth factors closely matched observed values, especially under haze and mist conditions. By adjusting $PM_{2.5}$ concentrations for RH based on visibility-derived growth factors, the accuracy of low-cost PM sensors was notably improved. This research highlights the importance of considering PM-meteorology interactions in visibility forecasting and sensor calibration.

In [156], a fuzzy combination forecasting system to enhance the accuracy and stability of air pollutant predictions is proposed. The system integrates data preprocessing, fuzzy theory, and a Cuckoo Search optimization algorithm to determine optimal model weights. Maximizing decorrelation and maintaining model diversity improves prediction reliability. Experiments using PM_{10} and $PM_{2.5}$ datasets from three cities demonstrate higher performance in accuracy, stability, and generalization, making the system highly effective for air quality early-warning applications, compared to state-of-the-art methods.

The paper [157] discusses the interpretability of a machine learning model predicting NO_2 concentration in Madrid. The Shapley Additive Explanations method is used to analyze the impact of individual features on the neural network predictions. Three different interpretation methods are also compared to determine their suitability for air quality data analysis. The results provide deeper insight into the model's performance and allow for a better understanding of its decisions.

In [158], an extended attention model for sequence-to-sequence recurrent neural networks designed to capture periodic patterns in time series is introduced. When applied to univariate and multivariate time series forecasting, the model demonstrates state-of-the-art performance. The model can be integrated with any RNN architecture, enhancing its ability to handle time series with underlying periodicity, thus improving forecasting accuracy.

The work [159] compares 10 state-of-the-art quantile regression models, focusing on forecasting up to 60 h ahead for forecasting NO_2 concentration. Instead of using raw quantiles, the study derives distribution parameters, making the approach semi-parametric. Quantile gradient-boosted trees perform best for both expected values and full distributions. However, a simpler quantile k-nearest neighbors model combined with linear regression achieves comparable results with lower computational costs. The findings highlight the advantages of probabilistic forecasting in air quality management, aiding authorities in activating traffic restrictions proactively.

The study [160] proposes a predictive model for SO_2 emissions, aiming to estimate high-emission episodes in advance. Unlike previous studies focusing on mean values, this approach uses quantile curves from additive models to capture the full distribution of pollution levels. A back-fitting algorithm with local polynomial kernel smoothers was applied, with bootstrapping used for hypothesis testing. The model was validated using simulated and real SO_2 data from a coal-fired power station in northern Spain, demonstrating its efficacy in forecasting pollution incidents.

The study [161] analyzes SO_2 air concentrations in Krakow (Poland) using data from two monitoring stations. The research identifies trends, seasonal variations, and cyclic patterns, applying a multiplicative time series model. Despite a slight long-term decline, an initial upward trend was observed. The study also examines wind direction effects, finding no significant impact, suggesting local emissions as the dominant pollution source. The methodology, validated through monthly and quarterly forecasts, is adaptable for analyzing other air pollutants in different cities.

Numerical chemical transport models, such as WRF-CHEM [162] and WRF-CMAQ [163], are widely used for simulating atmospheric pollutant dispersion and chemical transformations. These models provide high-resolution spatiotemporal forecasts, supporting both short-term nowcasting of pollution peaks and longer-term scenario analyses. They can be coupled with ground-based, satellite, or UAV measurements to improve forecast accuracy and enable comprehensive environmental assessments.

Discussion

The methods presented in this section illustrate the diversity of approaches applied to air pollution forecasting beyond conventional statistical, machine learning, or hybrid models. Deterministic chemical transport models, statistical time series methods, and novel machine learning or metaheuristic techniques each offer unique strengths, such as physical interpretability, data efficiency, or the ability to capture complex nonlinear relationships. Their application highlights the importance of tailoring the forecasting approach to the pollutant type, data availability, and spatial-temporal context.

Despite their potential, these approaches also have limitations. Deterministic models require extensive computational resources and high-quality input data, while statistical models may struggle with nonlinearities and transferability to new regions. Advanced ML and hybrid methods often improve predictive performance but can be complex to implement, computationally demanding, and difficult to interpret for policymakers.

Future research should focus on combining the strengths of different approaches, for example, by integrating physical models with adaptive machine learning or by using probabilistic and interpretability-focused techniques. Such integration could enhance forecast robustness, account for uncertainty, and provide actionable insights for regulatory planning, early warning systems, and public health interventions.

Overall, these methods complement mainstream forecasting approaches, demonstrating that diverse, context-sensitive strategies are essential for accurately predicting air pollutant concentrations and supporting effective environmental decision-making.

8. Summary, Open Directions, and Future Work

This article presents a comprehensive narrative review of air pollution forecasting methods, including traditional statistical approaches, modern techniques based on machine learning and deep learning, and hybrid models that combine different paradigms to improve prediction accuracy.

Despite significant technological progress, current forecasting methods still face limitations that affect their effectiveness and practical applicability. Many models insufficiently consider dynamic changes in emission sources, such as industrial accidents, wildfires, or traffic surges, and they often rely solely on historical data, limiting adaptability to new environmental conditions. In addition, forecasts frequently lack fine spatial resolution and do not use micro-scale meteorological phenomena. Moreover, few approaches explicitly account for the long-term impacts of climate change on pollutant dispersion and accumulation. Another important challenge remains the limited interpretability of advanced machine learning models, particularly deep neural networks.

Table 7 summarizes the main categories of data-driven forecasting methods reviewed in this paper, highlighting their typical techniques, strengths, and limitations. This synthesis supports the critical comparison of approaches and provides context for the discussion of future research directions.

Table 7. Comparison of forecasting method categories with selected algorithms/methods.

Category	Typical Methods	Strengths	Limitations
Deep Learning	Autoencoders, ANN, MLP, CNN, LSTM, Bi-LSTM, GRU, RNN, GCN, RBF, Transformer-based models	Captures complex non-linear spatiotemporal patterns; high predictive accuracy; adaptable to multi-modal data; automated feature extraction; supports transfer learning	"Black box" nature; requires large datasets and computational resources; prone to overfitting; location-specific tuning often needed
Machine Learning	SVM, RF, XGBoost, GAM, Prophet, *k*-NN, DT, GB	Handles nonlinearities; moderate interpretability; works with smaller datasets than DL; flexible for various features	Requires manual feature engineering; performance depends on data quality; may struggle with dynamic spatiotemporal patterns; limited generalization across regions
Statistical	AR, ARMA, ARIMA, MLR, SARIMA, QR, VARMA	Simple and interpretable; low computational cost; well-suited for stationary or seasonal patterns; provides uncertainty estimates	Limited to linear relationships; poor spatial modeling; assumes stationarity in many cases
Hybrid	ARIMA+ANN, CNN+LSTM, GCN+LSTM, PSO+LSTM, ANFIS, SSA+LSTM	Combines strengths of multiple paradigms; models physical processes and nonlinear dynamics; integrates heterogeneous data sources	High complexity; requires expertise in multiple paradigms; may inherit limitations of components; higher risk of overfitting

Open directions and future work

The field of air pollution forecasting is undergoing a rapid transformation due to advances in artificial intelligence, sensing technologies, and computational infrastructure. Among the most promising developments are novel AI architectures. Transformer-based models and attention mechanisms are emerging as powerful tools for capturing long-range temporal dependencies and multi-scale patterns in air quality data. Similarly, graph-based neural architectures allow for explicit representation of spatial relationships between monitoring stations, enabling improved predictions in geographically diverse regions.

Aircraft-based platforms, including drones, are increasingly employed to enhance air quality monitoring. These systems allow for high-resolution measurements in regions with complex terrain, industrial areas, or urban hotspots, providing complementary data that can be integrated with ground-based measurements and numerical models. Utilizing these synergistic approaches can improve the temporal and spatial resolution of pollutant retrieval and enhance forecasting accuracy, particularly in areas where traditional monitoring networks are sparse. Such integration supports more robust and actionable predictions for environmental management and policy decisions [164,165]. Despite these advantages, several open issues remain. Operational costs, regulatory restrictions, and the need for specialized personnel can limit widespread adoption. Additionally, data integration from heterogeneous sources, calibration of sensors, and handling large datasets in real time are ongoing challenges that require further research and development.

Recent large-scale satellite missions have demonstrated the practical integration of remote sensing, numerical modeling, and in situ measurements for comprehensive air quality retrieval and forecasting. For instance, the GEMS mission in Korea [166] and TEMPO in the United States [167] provide hourly observations of key pollutants across wide geographic regions, enabling near-real-time monitoring and policy support. Similarly, NASA missions such as Aura [168] and Sentinel-5P [169] have been extensively used to synergize satellite data with ground-based networks and chemical transport models, improving the spatial and temporal resolution of pollutant estimates. The Suomi National Polar-orbiting Partnership (Suomi NPP) mission, carrying instruments such as VIIRS and CrIS, provides critical observations of aerosols, clouds, and meteorological parameters that support both air quality forecasting and climate research [170]. NISAR, a joint NASA and ISRO mission, is designed to monitor changes in Earth's surface, including glacial melting, ground movements, and seismic activity [171]; data from this mission can also inform assessments of how environmental dynamics affect air quality. These initiatives exemplify how multi-source data fusion can enhance large-scale environmental assessments, offering valuable insights for air quality management. Future work may utilize these datasets in combination with local low-cost sensors and machine learning algorithms to develop more accurate, scalable, and robust forecasting frameworks.

Numerical chemical transport models, such as WRF-CHEM [162] and WRF-CMAQ [163], are widely used for simulating atmospheric pollutant dispersion and chemical transformations. These models provide high-resolution spatiotemporal forecasts, supporting both short-term nowcasting of pollution peaks and longer-term scenario analyses. They can be coupled with ground-based, satellite, or UAV measurements to improve forecast accuracy and enable comprehensive environmental assessments. For example, in [172], satellite-derived NO_2 columns over southern China were combined with meteorological outputs from the WRF model to produce high-resolution retrievals. The study demonstrated that integrating WRF data improves spatial accuracy, particularly in regions with complex terrain and large variability, and enables effective validation with ground-based MAX-DOAS measurements. Similarly, in [173], WRF-Chem simulations over southeastern Brazil were evaluated against satellite and ground-based observations for $PM_{2.5}$, NO_x, CO, O_3, and aerosol optical depth (AOD). The work highlighted that combining numerical modeling, remote sensing, and ground measurements allows accurate reproduction of pollutant patterns, captures the transport of urban emissions to inland areas, and underscores the value of integrated approaches for regional air quality assessment. Moreover, in a real-time forecasting context, WRF-Chem was used to support flight planning during the NOAA AEROMMA and NASA STAQS 2023 field campaigns [174]. The forecasting system provided 72 h predictions of atmospheric composition, aerosols, and clouds over domains centered on Chicago and New York City. Evaluation against lidar and surface observations showed that while forecast skill decreased with lead time, the system successfully captured the influence of transported pollution, including wildfire smoke from Canada, and identified limitations related to local meteorological simulations. This demonstrates the practical utility of WRF-Chem for operational forecasting and campaign planning, complementing the long-term modeling and retrieval studies discussed above. Recent studies in Southeast Asia have further illustrated the power of combining WRF-Chem with advanced satellite observations. For instance, the launch of the geostationary UV-VIS spectrometer GEMS in 2020 and NASA's ASIA-AQ mission in early 2024 enabled detailed monitoring of NO_2 over densely populated regions, including the Seoul Metropolitan Area, Manila, Taiwan, South Korea, and Thailand [175]. High-resolution WRF-Chem simulations

(4 km grid) revealed strong sensitivity of NO_2 columns to model resolution, captured local enhancements not resolved at coarser scales, and highlighted biases in global emission inventories (e.g., EDGAR v5). Integration with in situ airborne measurements further validated model outputs, demonstrating the potential of synergizing WRF-Chem with satellite and field campaign data for accurate, high-resolution air quality assessment and emissions adjustment in rapidly evolving environments.

Beyond technological advancements, the accessibility and openness of air quality data play a crucial role in enabling informed decision-making, fostering citizen engagement, and supporting smart city initiatives. Platforms such as OpenAQ provide raw air quality measurements that can be analyzed locally and globally, yet challenges remain regarding data quality, interoperability, and standardization across cities. Citizen science initiatives, like "Smell Pittsburgh" [176] and "Participatory Action for Citizens' Engagement" [177], demonstrate how community involvement in data collection and analysis can improve understanding of local air pollution issues and drive effective interventions. Future efforts combining low-cost sensors, satellite data, numerical models, and active public engagement have the potential to create more robust, equitable, and actionable air quality monitoring systems [178–180].

Another important direction is the use of federated and privacy-preserving learning, which allows models to be trained collaboratively across multiple institutions or regions without sharing raw data. This approach can address legal and ethical concerns related to privacy, regulatory restrictions, and data sovereignty. In parallel, there is growing interest in multi-modal and multi-fidelity data fusion. By combining ground-based sensor measurements, satellite imagery, chemical transport model outputs, meteorological forecasts, and mobility data, researchers can construct richer and more resilient forecasting frameworks.

Interpretability remains a major challenge in the deployment of deep learning models. Many state-of-the-art systems still operate as "black boxes", which can limit their acceptance by environmental agencies and policymakers. Alongside interpretability, there is a push toward real-time and edge computing solutions. Deploying forecasting algorithms directly on IoT devices or edge servers reduces latency and dependence on centralized infrastructure, enabling on-site air quality warnings in smart city applications.

Climate change introduces additional complexity to air pollution forecasting. Shifts in atmospheric circulation, the increasing frequency of wildfires, and the occurrence of heatwaves are altering pollution baselines and generating extreme events that existing models may not capture effectively. Utilizing climate-related variables in forecasting frameworks could improve resilience against these emerging challenges. Finally, the integration of forecasting systems into policy making and public health decision support tools remains an open problem. Models that produce actionable alerts, guide traffic or industrial restrictions, and inform urban planning can transform air quality forecasting from an academic exercise into a cornerstone of environmental governance. Key future research directions are as follows:

- Federated and privacy-preserving learning for multi-region, data-sensitive applications;
- Multi-modal data fusion integrating sensors, satellites, chemical transport models, and mobility data;
- Edge computing for real-time forecasting and on-site air quality alerts;
- Explainable AI to improve adoption by policymakers and environmental agencies;
- Adaptive hybrid models capable of responding to dynamic atmospheric, emission, and socio-economic changes;
- Integration of forecasting systems into public health and urban planning decision support tools;
- Development of standardized benchmarks and open datasets for reproducible evaluation.

The progress of the field critically depends on stronger benchmarking practices, as the current lack of standardized datasets, common evaluation metrics, and open repositories makes objective model comparison and reproducibility challenging. Many studies still rely on proprietary or locally collected data, which hinders cross-study evaluation and slows the adoption of best practices. Establishing open, curated benchmark datasets covering diverse regions and pollution scenarios, together with harmonized evaluation protocols, would facilitate fair model comparison, accelerate methodological innovation, and strengthen transparency. Achieving these goals will require interdisciplinary collaboration between environmental scientists, computer scientists, and policymakers, ensuring that advances in forecasting methods are both scientifically rigorous and practically applicable in real-world air quality management systems.

Future research should focus on increasing the transparency of models and developing methods that will allow for a better understanding of the impact of individual factors on air quality. It will be important to develop more adaptive algorithms that automatically adjust to new conditions, such as the introduction of emission regulations or changes in the transport structure. It will also be crucial to develop short-term forecasts that will enable a faster response to smog episodes and warn residents about threats. Moreover, further improvement of the models by integrating different data sources and developing hybrid forecasting methods should be realized. It will be necessary to develop more advanced methods of air quality forecasting that take into account variable atmospheric conditions, such as wind speed, wind direction, humidity, and pressure. These factors have a significant impact on the spread of pollutants, and their accurate modeling can significantly improve the precision of forecasts. Integrating meteorological data will allow for better prediction of sudden increases in pollution and their local effects. The development of such methods will help to prevent smog episodes more effectively and better plan protective measures. Therefore, future research should focus on creating models that take into account both historical data and dynamic changes in atmospheric conditions.

Author Contributions: Conceptualization, J.B.; methodology, J.B.; validation, J.B., investigation, J.B.; resources, R.S.; data curation, J.B.; writing—original draft preparation, J.B.; writing—review and editing, J.B.; visualization, J.B.; supervision, R.S.; funding acquisition, R.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Appendix A.1 Abbreviations of Methods Used for Forecasting

The abbreviations of methods cited in this survey are listed in Table A1.

Table A1. Abbreviations of methods used for forecasting.

Abbreviation	Method
ABC	artificial bee colony
AI	artificial intelligence
ANFIS	adaptive neuro-fuzzy inference system
ANN	artificial neural network
AOD	aerosol optical depth
API	Air Pollution Index
AR	autoregressive model
ARIMA	autoregressive integrated moving average
ARMA	autoregressive moving average
AT-GCN	attention temporal graph convolutional network
AQI	Air Quality Index
AQHI	Air Quality Health Index
Bi-GRU	Bi-directional gated recurrent unit
Bi-LSTM	Bi-directional long short-term memory
CAI	Comprehensive Air-quality Index
CAQI	Common Air Quality Index
CMAQ	Community Multiscale Air Quality
CNN	convolutional neural network
DAQI	Daily Air Quality Index
DL	deep learning model
DT	decision tree
DWT	discrete wavelet transform
EEA	European Environment Agency
EL	ensemble learning
ELM	extreme learning machine
EMD	empirical mode decomposition
ENN	Elman neural network
ESN	echo state networks
FAA-NN	factor-augmented autoregressive neural networks
FFNN	feed-forward neural network
GA	genetic algorithm
GAGNN	group-aware graph neural network
GAM	general additive model
GB	gradient-boosting-based machine learning
GBR	gradient boosting regression
GBTR	gradient boosted tree regression
GCN	graph convolutional network
GNN	graph neural network
GPR	Gaussian process regression
GRU	gated recurrent unit
HGN	hierarchical graph network
IAQI	Individual Air Quality Index
IoT	Internet of Things
IVS	input variable selection
k-NN	*k*-nearest neighbors
k-NNR	*k*-nearest neighbors regression
LR	linear regression
LSTM	long short-term memory
MVO	multi-verse optimization algorithm
MLP	multilayer perceptron
MLR	multiple linear regression
MTL	multi-task learning
NAR-NN	nonlinear autoregressive neural network
NARX	nonlinear autoregressive network with exogenous inputs
NOM	numerical optimization methods
PCA	principal component analysis
PSI	Pollutant Standards Index
PSO	particle swarm optimization
RBF	radial basis function
RBN	radial basis network
RF	random forest
RFR	random forest regression

Table A1. *Cont.*

Abbreviation	Method
RM	regression models
RNN	recurrent neural network
RT	regression trees
SARIMA	seasonal ARIMA
SR	stochastic regression
SSA	sparrow search algorithm
SSH-GNN	self-supervised hierarchical graph neural network
STN	sparse attention transformer network
SVM	support vector machine
SVR	support vector regression
TSP	total suspended particle
UAV	unmanned aerial vehicle
QC	quantile curves
QR	quantile regression
VAE	variational autoencoder
VARMA	vector ARMA
WPD	wavelet packet decomposition
WT-ARIMA	wavelet transform ARIMA
WANN	wavelet ANN
WRF	Weather Research and Forecasting
WTNN	wavelet transform neural network

Appendix A.2 Error Measures

Let x_i be the actual value, y_i is a predicted value and n is the number of observations. The $\bar{y}$ denotes the mean value of y, and the σ is the standard deviation. The error measures used for forecasting evaluation are listed in Table A2.

Table A2. Error measures (with abbreviations) used for forecasting evaluation.

Abbreviation	Error Measure	Equation
ρ	Spearman correlation	(A1)
AAD	absolute average deviation	(A2)
ARV	average relative variance	(A3)
CE	cross entropy	(A4)
DA	direction accuracy	(A5)
E^2	coefficient of efficiency	(A6)
EM	error mean	(A7)
IA	index of agreement	(A8)
MAE	mean absolute error	(A9)
MAPE	mean absolute percentage error	(A10)
MASE	mean absolute scaled error	(A11)
MFB	mean fractional bias	(A12)
NAE	normalized absolute error	(A13)
NMSE	normalized mean square error	(A14)
NRMSE	normalized root mean square error	(A15)
PA	prediction accuracy	(A16)
R	Pearson correlation coefficient	(A17)
R^2	coefficient of determination	(A18)
RE	relative error	(A19)
RMS	root mean square	(A20)
RMSE	root mean square error	(A21)
SEP	standard error of prediction	(A22)
SMSE	standardized mean squared error	(A23)
SOS	sum-of-squares	(A24)
SSE	sum square error	(A25)

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{A1}$$

where $d_i = R(x_i) - R(y_i)$ is a rank difference for each pair (x_i, y_i), and $R(x_i)$, $R(y_i)$ – ranks of appropriate values.

$$\mathrm{AAD} = \frac{\sum_{i=1}^{n}(\frac{x_t - y_t}{x_t})}{n} \cdot 100\% \tag{A2}$$

$$\mathrm{ARV} = 1.0 - \mathrm{E}^2 \tag{A3}$$

$$\mathrm{CE} = -\sum_{i=1}^{n} x_i \ln\left(\frac{y_i}{x_i}\right) \tag{A4}$$

$$\mathrm{DA} = \frac{1}{n}\sum_{i=1}^{n} w_i, \quad w_i = \begin{cases} 1, & \text{if } (y_{i+1} - y_i)(x_{i+1} - y_i) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{A5}$$

$$\mathrm{E}^2 = 1.0 - \frac{\sum_{i=1}^{n} |x_i - y_i|^2}{\sum_{i=1}^{n} |x_i - \bar{y}|^2} \tag{A6}$$

$$\mathrm{EM} = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i) \tag{A7}$$

$$\mathrm{IA} = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(|x_i - \bar{y}| + |y_i - \bar{y}|)^2} \tag{A8}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - x_i}{n}\right| \tag{A9}$$

$$\mathrm{MAPE} = 100\frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - y_i}{x_i}\right| \tag{A10}$$

$$\mathrm{MASE} = \frac{\frac{1}{n}\sum_{i=1}^{n} |x_i - y_i|}{\frac{1}{n-1}\sum_{j=1}^{n-1} |x_j - x_{j-1}|} \tag{A11}$$

$$\mathrm{MFB} = \frac{1}{n}\sum_{i=1}^{n} \frac{x_i - y_i}{\frac{1}{2}(x_i + y_i)} \tag{A12}$$

$$\mathrm{NAE} = \frac{\sum_{i=1}^{n} |x_i - y_i|}{\sum_{i=1}^{n} y_i} \tag{A13}$$

$$\mathrm{NMSE} = \frac{\sum_{i=1}^{n} \overline{(y_i - x_i)^2}}{\sum_{i=1}^{n} \overline{y_i x_i}} \tag{A14}$$

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{x_{\max} - x_{\min}} \tag{A15}$$

$$\mathrm{PA} = \sum_{i=1}^{n} \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{(n-1)\sigma(x_i)\sigma(y_i)} \tag{A16}$$

$$\mathrm{R} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{A17}$$

$$\mathrm{R}^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{\sum_{i=1}^{n} (x_i - \bar{y})^2} \tag{A18}$$

$$\mathrm{RE} = \frac{\sum_{i=1}^{n} |x_i - y_i|}{\sum_{i=1}^{n} x_i} \cdot 100\% \tag{A19}$$

$$\mathrm{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} x_i^2} \tag{A20}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(x_i - y_i\right)^2} \tag{A21}$$

$$\mathrm{SEP} = \frac{100}{\bar{x}} \mathrm{RMSE} \tag{A22}$$

$$\mathrm{SMSE} = \frac{1}{n} \frac{\sum_{i=1}^{n} (x_i - y_i)^2}{\sigma_y^2} \tag{A23}$$

$$\mathrm{SOS} = \sum_{i=1}^{n} (y_i - x_i)^2 \tag{A24}$$

$$\mathrm{SSE} = \sum_{i=1}^{n} (x_i - y_i)^2 \tag{A25}$$

References

1. Wu, J.Z.; Ge, D.D.; Zhou, L.F.; Hou, L.Y.; Zhou, Y.; Li, Q.Y. Effects of particulate matter on allergic respiratory diseases. *Chronic Dis. Transl. Med.* **2018**, *4*, 95–102. [CrossRef]
2. Mak, H.W.L.; Ng, D.C.Y. Spatial and Socio-Classification of Traffic Pollutant Emissions and Associated Mortality Rates in High-Density Hong Kong via Improved Data Analytic Approaches. *Int. J. Environ. Res. Public Health* **2021**, *18*, 6532. [CrossRef]
3. Chen, F.; Zhang, W.; Mfarrej, M.F.B.; Saleem, M.H.; Khan, K.A.; Ma, J.; Raposo, A.; Han, H. Breathing in danger: Understanding the multifaceted impact of air pollution on health impacts. *Ecotoxicol. Environ. Saf.* **2024**, *280*, 116532. [CrossRef]
4. Hamra, G.B.; Guha, N.; Cohen, A.; Laden, F.; Raaschou-Nielsen, O.; Samet, J.M.; Vineis, P.; Forastiere, F.; Saldiva, P.; Yorifuji, T.; et al. Outdoor Particulate Matter Exposure and Lung Cancer: A Systematic Review and Meta-Analysis. *Environ. Health Perspect.* **2014**, *122*, 906–911. [CrossRef]
5. Połednik, B. Emissions of air pollution in industrial and rural region in Poland and health impacts. *J. Ecol. Eng.* **2022**, *23*, 250–258. [CrossRef]

6. Jo, E.J.; Lee, W.S.; Jo, H.Y.; Kim, C.H.; Eom, J.S.; Mok, J.H.; Kim, M.H.; Lee, K.; Kim, K.U.; Lee, M.K.; et al. Effects of particulate matter on respiratory disease and the impact of meteorological factors in Busan, Korea. *Respir. Med.* **2017**, *124*, 79–87. [CrossRef]
7. Seo, Y.; Kim, S.M. Estimation of greenhouse gas emissions from road traffic: A case study in Korea. *Renew. Sustain. Energy Rev.* **2013**, *28*, 777–787. [CrossRef]
8. Cheng, W.L.; Chen, Y.S.; Zhang, J.; Lyons, T.; Pai, J.L.; Chang, S.H. Comparison of the revised air quality index with the PSI and AQI indices. *Sci. Total Environ.* **2007**, *382*, 191–198. [CrossRef]
9. Swamee, P.K.; Tyagi, A. Formation of an air pollution index. *J. Air Waste Manag. Assoc.* **1999**, *49*, 88–91. [CrossRef]
10. Tang, K.T.J.; Lin, C.; Wang, Z.; Pang, S.W.; Wong, T.W.; Yu, I.T.S.; Fung, W.W.Y.; Hossain, M.S.; Lau, A.K. Update of Air Quality Health Index (AQHI) and harmonization of health protection and climate mitigation. *Atmos. Environ.* **2024**, *326*, 120473. [CrossRef]
11. Wang, B.; Yi, D.; Jin, L.; Li, Z.; Liu, J.; Zhang, Y.; Qiu, X.; Liu, W.; Tao, S.; Ren, A. Organochlorine pesticide levels in maternal serum and risk of neural tube defects in offspring in Shanxi Province, China: A case–control study. *Sci. Total Environ.* **2014**, *490*, 1037–1043. [CrossRef]
12. Wang, Z.; Zhang, R.; Chen, R.; Chen, H. The Aerosol Optical Depth Retrieval from Wide-Swath Imaging of DaQi-1 over Beijing. *Atmosphere* **2024**, *15*, 1476. [CrossRef]
13. Lee, J.Y.; Jo, W.K.; Chun, H.H. Long-term trends in visibility and its relationship with mortality, air-quality index, and meteorological factors in selected areas of Korea. *Aerosol Air Qual. Res.* **2015**, *15*, 673–681. [CrossRef]
14. Chen, W.; Tang, H.; Zhao, H. Urban air quality evaluations under two versions of the national ambient air quality standards of China. *Atmos. Pollut. Res.* **2016**, *7*, 49–57. [CrossRef]
15. Wang, X.; Zhang, S.; Chen, Y.; He, L.; Ren, Y.; Zhang, Z.; Li, J.; Zhang, S. Air quality forecasting using a spatiotemporal hybrid deep learning model based on VMD–GAT–BiLSTM. *Sci. Rep.* **2024**, *14*, 17841. [CrossRef]
16. Subramaniam, S.; Raju, N.; Ganesan, A.; Rajavel, N.; Chenniappan, M.; Prakash, C.; Pramanik, A.; Basak, A.K.; Dixit, S. Artificial Intelligence Technologies for Forecasting Air Pollution and Human Health: A Narrative Review. *Sustainability* **2022**, *14*, 9951. [CrossRef]
17. Zafra-Pérez, A.; Medina-García, J.; Boente, C.; Gómez-Galán, J.; Sánchez de la Campa, A.; de la Rosa, J. Designing a low-cost wireless sensor network for particulate matter monitoring: Implementation, calibration, and field-test. *Atmos. Pollut. Res.* **2024**, *15*, 102208. [CrossRef]
18. Apostolopoulos, I.D.; Fouskas, G.; Pandis, S.N. Field Calibration of a Low-Cost Air Quality Monitoring Device in an Urban Background Site Using Machine Learning Models. *Atmosphere* **2023**, *14*, 368. [CrossRef]
19. Pochwała, S.; Gardecki, A.; Lewandowski, P.; Somogyi, V.; Anweiler, S. Developing of Low-Cost Air Pollution Sensor—Measurements with the Unmanned Aerial Vehicles in Poland. *Sensors* **2020**, *20*, 3582. [CrossRef]
20. Jaroń, A.; Borucka, A.; Deliś, P.; Sekrecka, A. An Assessment of the Possibility of Using Unmanned Aerial Vehicles to Identify and Map Air Pollution from Infrastructure Emissions. *Energies* **2024**, *17*, 577. [CrossRef]
21. Johnson, T.; Woodward, K. Enviro-IoT: Calibrating Low-Cost Environmental Sensors in Urban Settings. *arXiv* **2025**, arXiv:2502.07596.
22. Moskal, A.; Jagodowicz, W.; Penconek, A.; Zaraska, K. Low-Cost Sensor System for Air Purification Process Evaluation. *Sensors* **2024**, *24*, 1769. [CrossRef]
23. De Vito, S.; D'Elia, G.; Ferlito, S.; Francia, G.D.; Davidović, M.D.; Kleut, D.; Stojanović, D.; Jovaševic-Stojanović, M. A Global Multiunit Calibration as a Method for Large-Scale IoT Particulate Matter Monitoring Systems Deployments. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 2501916. [CrossRef]
24. Kumar, G.; D, P.K.; Dhariwal, J.; Srirangarajan, S. Reliability Assessment of Low-Cost PM Sensors under High Humidity and High PM Level Outdoor Conditions. *IEEE Sens. J.* **2025**, *25*, 32746–32758. [CrossRef]
25. Ken, H.M.; Behjati, M.; Rafsanjani, A.S.; Aslam, S.; Meng, Y.K.; Majeed, A.P.P.A.; Zheng, Y. Advancing Air Quality Monitoring: TinyML-Based Real-Time Ozone Prediction with Cost-Effective Edge Devices. In *Selected Proceedings from the 2nd International Conference on Intelligent Manufacturing and Robotics, ICIMR 2024, 22–23 August, Suzhou, China*; Springer Nature: Singapore, 2025; pp. 502–512. [CrossRef]
26. Yadav, K.; Arora, V.; Jha, S.K.; Kumar, M.; Tripathi, S.N. Few-shot calibration of low-cost air pollution (PM2.5) sensors using meta-learning. *IEEE Sens. Lett.* **2021**, *6*, 1–4. [CrossRef]
27. Xu, R.; Yao, D.; Pian, Y.; Cao, R.; Fu, Y.; Yang, X.; Gan, T.; Liu, Y. Integrating mobile and fixed monitoring data for high-resolution PM2.5 mapping using machine learning. *arXiv* **2025**, arXiv:2503.12367. [CrossRef]
28. Wang, G.; Yu, C.; Guo, K.; Guo, H.; Wang, Y. Research of low-cost air quality monitoring models with different machine learning algorithms. *Atmos. Meas. Tech.* **2024**, *17*, 181–196. [CrossRef]
29. Parra-Medina, F.D.; Vélez-Guerrero, M.A.; Callejas-Cuervo, M. Low-Cost Solution for Air Quality Monitoring: Unmanned Aerial System and Data Transmission via LoRa Protocol. *Sustainability* **2024**, *16*, 10108. [CrossRef]

30. Ahmadi, M.H.; Jashnani, H.; Chau, K.W.; Kumar, R.; Rosen, M.A. Carbon dioxide emissions prediction of five Middle Eastern countries using artificial neural networks. *Energy Sources Part A Recover. Util. Environ. Eff.* **2023**, *45*, 9513–9525. [CrossRef]
31. Cho, J.H.; Moon, J.W. Integrated artificial neural network prediction model of indoor environmental quality in a school building. *J. Clean. Prod.* **2022**, *344*, 131083. [CrossRef]
32. Sohn, S.H.; Oh, S.C.; Yeo, Y.K. Prediction of air pollutants by using an artificial neural network. *Korean J. Chem. Eng.* **1999**, *16*, 382–387. [CrossRef]
33. Liang, Y.; Ma, J.; Tang, C.; Ke, N.; Wang, D. Hourly forecasting on PM2.5 concentrations using a deep neural network with meteorology inputs. *Environ. Monit. Assess.* **2023**, *195*, 1510. [CrossRef]
34. He, Z.; Guo, Q.; Wang, Z.; Li, X. Prediction of monthly PM2.5 concentration in Liaocheng in China employing artificial neural network. *Atmosphere* **2022**, *13*, 1221. [CrossRef]
35. Guo, Q.; He, Z.; Wang, Z. Prediction of Hourly PM2.5 and PM10 Concentrations in Chongqing City in China Based on Artificial Neural Network. *Aerosol Air Qual. Res.* **2023**, *23*, 220448. [CrossRef]
36. Guo, Q.; He, Z.; Wang, Z. Predicting of daily PM2.5 concentration employing wavelet artificial neural networks based on meteorological elements in Shanghai, China. *Toxics* **2023**, *11*, 51. [CrossRef]
37. Gogikar, P.; Tyagi, B.; Gorai, A. Seasonal prediction of particulate matter over the steel city of India using neural network models. *Model. Earth Syst. Environ.* **2019**, *5*, 227–243. [CrossRef]
38. Gallo, C.; Conto, F.; Fiore, M. A neural network model for forecasting CO_2 emission. *AGRIS On-Line Pap. Econ. Inform.* **2014**, *6*, 31–36.
39. Ramentol, E.; Grimm, S.; Stinzendörfer, M.; Wagner, A. Short-Term Air Pollution Forecasting Using Embeddings in Neural Networks. *Atmosphere* **2023**, *14*, 298. [CrossRef]
40. Chae, S.; Shin, J.; Kwon, S.; Lee, S.; Kang, S.; Lee, D. PM10 and PM2.5 real-time prediction models using an interpolated convolutional neural network. *Sci. Rep.* **2021**, *11*, 11952. [CrossRef]
41. Du, S.; Li, T.; Yang, Y.; Horng, S.J. Deep air quality forecasting using hybrid deep learning framework. *IEEE Trans. Knowl. Data Eng.* **2019**, *33*, 2412–2424. [CrossRef]
42. Bekkar, A.; Hssina, B.; Douzi, S.; Douzi, K. Air-pollution prediction in smart city, deep learning approach. *J. Big Data* **2021**, *8*, 161. [CrossRef]
43. Gilik, A.; Ogrenci, A.S.; Ozmen, A. Air quality prediction using CNN+ LSTM-based hybrid deep learning architecture. *Environ. Sci. Pollut. Res.* **2022**, *29*, 11920–11938. [CrossRef]
44. Pak, U.; Ma, J.; Ryu, U.; Ryom, K.; Juhyok, U.; Pak, K.; Pak, C. Deep learning-based PM2.5 prediction considering the spatiotemporal correlations: A case study of Beijing, China. *Sci. Total Environ.* **2020**, *699*, 133561. [CrossRef]
45. Sharma, E.; Deo, R.C.; Prasad, R.; Parisi, A.V.; Raj, N. Deep air quality forecasts: Suspended particulate matter modeling with convolutional neural and long short-term memory networks. *IEEE Access* **2020**, *8*, 209503–209516. [CrossRef]
46. Li, T.; Hua, M.; Wu, X. A hybrid CNN-LSTM model for forecasting particulate matter (PM2.5). *IEEE Access* **2020**, *8*, 26933–26940. [CrossRef]
47. Liu, D.R.; Lee, S.J.; Huang, Y.; Chiu, C.J. Air pollution forecasting based on attention-based LSTM neural network and ensemble learning. *Expert Syst.* **2020**, *37*, e12511. [CrossRef]
48. Dairi, A.; Harrou, F.; Khadraoui, S.; Sun, Y. Integrated Multiple Directed Attention-Based Deep Learning for Improved Air Pollution Forecasting. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3520815. [CrossRef]
49. Mengara Mengara, A.G.; Park, E.; Jang, J.; Yoo, Y. Attention-Based Distributed Deep Learning Model for Air Quality Forecasting. *Sustainability* **2022**, *14*, 3269. [CrossRef]
50. Liu, B.; Yan, S.; Li, J.; Qu, G.; Li, Y.; Lang, J.; Gu, R. An Attention-Based Air Quality Forecasting Method. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 728–733. [CrossRef]
51. Elbaz, K.; Shaban, W.M.; Zhou, A.; Shen, S.L. Real time image-based air quality forecasts using a 3D-CNN approach with an attention mechanism. *Chemosphere* **2023**, *333*, 138867. [CrossRef]
52. Abdullah, S.; Ismail, M.; Ahmed, A.N.; Abdullah, A.M. Forecasting particulate matter concentration using linear and non-linear approaches for air quality decision support. *Atmosphere* **2019**, *10*, 667. [CrossRef]
53. Gradišar, D.; Grašič, B.; Božnar, M.Z.; Mlakar, P.; Kocijan, J. Improving of local ozone forecasting by integrated models. *Environ. Sci. Pollut. Res.* **2016**, *23*, 18439–18450. [CrossRef] [PubMed]
54. Shams, S.R.; Jahani, A.; Kalantary, S.; Moeinaddini, M.; Khorasani, N. The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models for predicting SO_2 concentration. *Urban Clim.* **2021**, *37*, 100837. [CrossRef]
55. Kurnaz, G.; Demir, A.S. Prediction of SO_2 and PM10 air pollutants using a deep learning-based recurrent neural network: Case of industrial city Sakarya. *Urban Clim.* **2022**, *41*, 101051. [CrossRef]
56. Chang, Y.S.; Chiao, H.T.; Abimannan, S.; Huang, Y.P.; Tsai, Y.T.; Lin, K.M. An LSTM-based aggregated model for air pollution forecasting. *Atmos. Pollut. Res.* **2020**, *11*, 1451–1463. [CrossRef]

57. Song, C.; Fu, X. Research on different weight combination in air quality forecasting models. *J. Clean. Prod.* **2020**, *261*, 121169. [CrossRef]
58. Guo, Q.; He, Z.; Li, S.; Li, X.; Meng, J.; Hou, Z.; Liu, J.; Chen, Y. Air Pollution Forecasting Using Artificial and Wavelet Neural Networks with Meteorological Conditions. *Aerosol Air Qual. Res.* **2020**, *20*, 1429–1439. [CrossRef]
59. Banasiewicz, A.; Moosavi, F.; Kotyla, M.; Śliwiński, P.; Krot, P.; Wodecki, J.; Zimroz, R. Forecasting of NOx emissions of diesel LHD vehicles in underground mines—An ANN-based regression approach. *Appl. Sci.* **2023**, *13*, 9965. [CrossRef]
60. Fernández-Avilés, G.; Mattera, R.; Scepi, G. Factor-Augmented Autoregressive Neural Network to forecast Nox in the city of Madrid. *Socio-Econ. Plan. Sci.* **2024**, *95*, 101958. [CrossRef]
61. Liu, B.; Yu, X.; Chen, J.; Wang, Q. Air pollution concentration forecasting based on wavelet transform and combined weighting forecasting model. *Atmos. Pollut. Res.* **2021**, *12*, 101144. [CrossRef]
62. Mao, W.; Jiao, L.; Wang, W.; Wang, J.; Tong, X.; Zhao, S. A hybrid integrated deep learning model for predicting various air pollutants. *GISci. Remote Sens.* **2021**, *58*, 1395–1412. [CrossRef]
63. Abirami, S.; Chitra, P. Regional air quality forecasting using spatiotemporal deep learning. *J. Clean. Prod.* **2021**, *283*, 125341. [CrossRef]
64. Hickman, S.H.M.; Griffiths, P.T.; Nowack, P.J.; Archibald, A.T. Short-term forecasting of ozone air pollution across Europe with transformers. *Environ. Data Sci.* **2023**, *2*, e43. [CrossRef]
65. Zhang, Z.; Zhang, S. Modeling air quality PM2.5 forecasting using deep sparse attention-based transformer networks. *Int. J. Environ. Sci. Technol.* **2023**, *20*, 13535–13550. [CrossRef]
66. Xu, J.; Chen, L.; Lv, M.; Zhan, C.; Chen, S.; Chang, J. HighAir: A Hierarchical Graph Neural Network-Based Air Quality Forecasting Method. *arXiv* **2021**, arXiv:2101.04264.
67. Iskandaryan, D.; Ramos, F.; Trilles, S. Graph Neural Network for Air Quality Prediction: A Case Study in Madrid. *IEEE Access* **2023**, *11*, 2729–2742. [CrossRef]
68. Oliveira Santos, V.; Costa Rocha, P.A.; Scott, J.; Van Griensven Thé, J.; Gharabaghi, B. Spatiotemporal Air Pollution Forecasting in Houston-TX: A Case Study for Ozone Using Deep Graph Neural Networks. *Atmosphere* **2023**, *14*, 308. [CrossRef]
69. Chen, L.; Xu, J.; Wu, B.; Huang, J. Group-Aware Graph Neural Network for Nationwide City Air Quality Forecasting. *ACM Trans. Knowl. Discov. Data* **2023**, *18*, 1–20. [CrossRef]
70. Terroso-Saenz, F.; Morales-García, J.; Muñoz, A. Nationwide Air Pollution Forecasting with Heterogeneous Graph Neural Networks. *ACM Trans. Intell. Syst. Technol.* **2024**, *15*, 1–19. [CrossRef]
71. Han, J.; Liu, H.; Xiong, H.; Yang, J. Semi-Supervised Air Quality Forecasting via Self-Supervised Hierarchical Graph Neural Network. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 5230–5243. [CrossRef]
72. Dua, R.D.; Madaan, D.M.; Mukherjee, P.M.; Lall, B.L. Real Time Attention Based Bidirectional Long Short-Term Memory Networks for Air Pollution Forecasting. In Proceedings of the 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), Newark, CA, USA, 4–9 April 2019; pp. 151–158. [CrossRef]
73. Murillo-Escobar, J.; Sepulveda-Suescun, J.; Correa, M.; Orrego-Metaute, D. Forecasting concentrations of air pollutants using support vector regression improved with particle swarm optimization: Case study in Aburrá Valley, Colombia. *Urban Clim.* **2019**, *29*, 100473. [CrossRef]
74. Wang, J.; Niu, T.; Wang, R. Research and application of an air quality early warning system based on a modified least squares support vector machine and a cloud model. *Int. J. Environ. Res. Public Health* **2017**, *14*, 249. [CrossRef] [PubMed]
75. Zhu, D.; Cai, C.; Yang, T.; Zhou, X. A machine learning approach for air quality prediction: Model regularization and optimization. *Big Data Cogn. Comput.* **2018**, *2*, 5. [CrossRef]
76. Yu, Y.; Li, H.; Sun, S.; Li, Y. PM2.5 concentration forecasting through a novel multi-scale ensemble learning approach considering intercity synergy. *Sustain. Cities Soc.* **2022**, *85*, 104049. [CrossRef]
77. Lee, M.; Lin, L.; Chen, C.Y.; Tsao, Y.; Yao, T.H.; Fei, M.H.; Fang, S.H. Forecasting air quality in Taiwan by using machine learning. *Sci. Rep.* **2020**, *10*, 4153. [CrossRef]
78. Eddine, S.S.; Drissi, L.B.; Mejjad, N.; Mabrouki, J.; Romanov, A.A. Machine learning models application for spatiotemporal patterns of particulate matter prediction and forecasting over Morocco in north of Africa. *Atmos. Pollut. Res.* **2024**, *15*, 102239. [CrossRef]
79. Shen, J.; Valagolam, D.; McCalla, S. Prophet forecasting model: A machine learning approach to predict the concentration of air pollutants (PM2.5, PM10, O_3, NO_2, SO_2, CO) in Seoul, South Korea. *PeerJ* **2020**, *8*, e9961. [CrossRef]
80. Kaur, G.; Kaur, H. Ensemble based J48 and random forest based C6H6 air pollution detection. *Int. J. Sci. Res. Comput. Sci. Eng.* **2018**, *6*, 41–50. [CrossRef]
81. Mak, H. Improved Remote Sensing Algorithms and Data Assimilation Approaches in Solving Environmental Retrieval Problems. Ph.D. Thesis, Hong Kong University of Science and Technology (Hong Kong), Hong Kong, China, 2019. [CrossRef]
82. Ravindiran, G.; Hayder, G.; Kanagarathinam, K.; Alagumalai, A.; Sonne, C. Air quality prediction by machine learning models: A predictive study on the indian coastal city of Visakhapatnam. *Chemosphere* **2023**, *338*, 139518. [CrossRef] [PubMed]

83. Chen, S.; Wang, J.Q.; Zhang, H.Y. A hybrid PSO-SVM model based on clustering algorithm for short-term atmospheric pollutant concentration forecasting. *Technol. Forecast. Soc. Change* **2019**, *146*, 41–54. [CrossRef]
84. Lei, T.M.T.; Ng, S.C.W.; Siu, S.W.I. Application of ANN, XGBoost, and Other ML Methods to Forecast Air Quality in Macau. *Sustainability* **2023**, *15*, 5341. [CrossRef]
85. Ma, J.; Yu, Z.; Qu, Y.; Xu, J.; Cao, Y. Application of the XGBoost machine learning method in PM2.5 prediction: A case study of Shanghai. *Aerosol Air Qual. Res.* **2020**, *20*, 128–138. [CrossRef]
86. Tang, D.; Zhan, Y.; Yang, F. A review of machine learning for modeling air quality: Overlooked but important issues. *Atmos. Res.* **2024**, *300*, 107261. [CrossRef]
87. Zhang, L.; Lin, J.; Qiu, R.; Hu, X.; Zhang, H.; Chen, Q.; Tan, H.; Lin, D.; Wang, J. Trend analysis and forecast of PM2.5 in Fuzhou, China using the ARIMA model. *Ecol. Indic.* **2018**, *95*, 702–710. [CrossRef]
88. Zhang, H.; Zhang, S.; Wang, P.; Qin, Y.; Wang, H. Forecasting of particulate matter time series using wavelet analysis and wavelet-ARMA/ARIMA model in Taiyuan, China. *J. Air Waste Manag. Assoc.* **2017**, *67*, 776–788. [CrossRef] [PubMed]
89. Bhatti, U.A.; Yan, Y.; Zhou, M.; Ali, S.; Hussain, A.; Qingsong, H.; Yu, Z.; Yuan, L. Time series analysis and forecasting of air pollution particulate matter (PM 2.5): An SARIMA and factor analysis approach. *IEEE Access* **2021**, *9*, 41019–41031. [CrossRef]
90. Aladağ, E. Forecasting of particulate matter with a hybrid ARIMA model based on wavelet transformation and seasonal adjustment. *Urban Clim.* **2021**, *39*, 100930. [CrossRef]
91. Afrin, S.; Islam, M.M.; Ahmed, T. A meteorology based particulate matter prediction model for megacity Dhaka. *Aerosol Air Qual. Res.* **2021**, *21*, 200371. [CrossRef]
92. Yun, J.; Woo, J. IoT-enabled particulate matter monitoring and forecasting method based on cluster analysis. *IEEE Internet Things J.* **2020**, *8*, 7380–7393. [CrossRef]
93. Aznarte, J.L. Probabilistic forecasting for extreme NO_2 pollution episodes. *Environ. Pollut.* **2017**, *229*, 321–328. [CrossRef]
94. Kumar, U.; Jain, V. ARIMA forecasting of ambient air pollutants (O_3, NO, NO_2 and CO). *Stoch. Environ. Res. Risk Assess.* **2010**, *24*, 751–760. [CrossRef]
95. Vlachogianni, A.; Kassomenos, P.; Karppinen, A.; Karakitsios, S.; Kukkonen, J. Evaluation of a multiple regression model for the forecasting of the concentrations of NOx and PM10 in Athens and Helsinki. *Sci. Total Environ.* **2011**, *409*, 1559–1571. [CrossRef] [PubMed]
96. Gao, W.; Xiao, T.; Zou, L.; Li, H.; Gu, S. Analysis and Prediction of Atmospheric Environmental Quality Based on the Autoregressive Integrated Moving Average Model (ARIMA Model) in Hunan Province, China. *Sustainability* **2024**, *16*, 8471. [CrossRef]
97. Armstrong, E.G. A hybrid model of problem-based learning. In *The Challenge of Problem-Based Learning*, 1st ed.; Routledge: London, UK, 1997.
98. Mishra, D.; Goyal, P. Development of artificial intelligence based NO_2 forecasting models at Taj Mahal, Agra. *Atmos. Pollut. Res.* **2015**, *6*, 99–106. [CrossRef]
99. Kujawska, J.; Kulisz, M.; Oleszczuk, P.; Cel, W. Machine learning methods to forecast the concentration of PM10 in Lublin, Poland. *Energies* **2022**, *15*, 6428. [CrossRef]
100. Gaj, K.; Miller, U.; Sówka, I. Progressing climate changes and deteriorating air quality as one of the biggest challenges of sustainable development of the present decade in Poland. *Sustainability* **2020**, *12*, 6367. [CrossRef]
101. Kryza, M.; Werner, M.; Guzikowski, J. Assimilation of Meteorological Data in Online Integrated Atmospheric Transport Model—Example of Air Quality Forecasts for Poland. In *International Technical Meeting on Air Pollution Modelling and its Application*; Springer: Cham, Switzerland, 2020; pp. 273–278.
102. Usha Ruby, A.; Chandran, J.G.C.; Theerthagiri, P.; Patil, R.; Chaithanya, B.; Jain, T.S. Forecasting PM2.5 Concentration Using Gradient-Boosted Regression Tree with CNN Learning Model. *Opt. Mem. Neural Netw.* **2024**, *33*, 86–96. [CrossRef]
103. Díaz-Robles, L.A.; Ortega, J.C.; Fu, J.S.; Reed, G.D.; Chow, J.C.; Watson, J.G.; Moncada-Herrera, J.A. A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmos. Environ.* **2008**, *42*, 8331–8340. [CrossRef]
104. Neto, P.S.D.M.; Firmino, P.R.A.; Siqueira, H.; Tadano, Y.D.S.; Alves, T.A.; De Oliveira, J.F.L.; Marinho, M.H.D.N.; Madeiro, F. Neural-based ensembles for particulate matter forecasting. *IEEE Access* **2021**, *9*, 14470–14490. [CrossRef]
105. Al Yammahi, A.; Aung, Z. Forecasting the concentration of NO_2 using statistical and machine learning methods: A case study in the UAE. *Heliyon* **2023**, *9*, e12584. [CrossRef]
106. Shams, S.R.; Jahani, A.; Kalantary, S.; Moeinaddini, M.; Khorasani, N. Artificial intelligence accuracy assessment in NO_2 concentration forecasting of metropolises air. *Sci. Rep.* **2021**, *11*, 1805. [CrossRef] [PubMed]
107. Wang, P.; Liu, Y.; Qin, Z.; Zhang, G. A novel hybrid forecasting model for PM10 and SO_2 daily concentrations. *Sci. Total Environ.* **2015**, *505*, 1202–1212. [CrossRef]

108. Sánchez, A.B.; Ordóñez, C.; Lasheras, F.S.; de Cos Juez, F.J.; Roca-Pardiñas, J. Forecasting SO_2 pollution incidents by means of Elman artificial neural networks and ARIMA models. In *Abstract and Applied Analysis*; Wiley Online Library: Hoboken, NJ, USA, 2013; Volume 2013, p. 238259.
109. Lee, S.; Kim, M. Application of Long-short Term Memory (LSTM) Model for Forecasting NOx Emission in Pohang Area. *arXiv* **2023**, arXiv:2311.15632. [CrossRef]
110. Dobrea, M.; Bădicu, A.; Barbu, M.; Subea, O.; Bălănescu, M.; Suciu, G.; Bîrdici, A.; Orza, O.; Dobre, C. Machine Learning algorithms for air pollutants forecasting. In Proceedings of the 2020 IEEE 26th International Symposium for Design and Technology in Electronic Packaging (SIITME), Pitesti, Romania, 21–24 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 109–113.
111. Sharma, M.; Jain, S.; Mittal, S.; Sheikh, T.H. Forecasting and prediction of air pollutants concentrates using machine learning techniques: The case of India. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2021; Volume 1022, p. 012123.
112. Liu, H.; Jin, K.; Duan, Z. Air PM2.5 concentration multi-step forecasting using a new hybrid modeling method: Comparing cases for four cities in China. *Atmos. Pollut. Res.* **2019**, *10*, 1588–1600. [CrossRef]
113. Zaini, N.A.; Ahmed, A.N.; Ean, L.W.; Chow, M.F.; Malek, M.A. Forecasting of fine particulate matter based on LSTM and optimization algorithm. *J. Clean. Prod.* **2023**, *427*, 139233. [CrossRef]
114. Dirik, M. Prediction of NOx emissions from gas turbines of a combined cycle power plant using an ANFIS model optimized by GA. *Fuel* **2022**, *321*, 124037. [CrossRef]
115. Heydari, A.; Majidi Nezhad, M.; Astiaso Garcia, D.; Keynia, F.; De Santoli, L. Air pollution forecasting application based on deep learning model and optimization algorithm. *Clean Technol. Environ. Policy* **2022**, *24*, 607–621. [CrossRef]
116. Udristioiu, M.T.; Mghouchi, Y.E.; Yildizhan, H. Prediction, modelling, and forecasting of PM and AQI using hybrid machine learning. *J. Clean. Prod.* **2023**, *421*, 138496. [CrossRef]
117. Juhos, I.; Makra, L.; Tóth, B. Forecasting of traffic origin NO and NO_2 concentrations by Support Vector Machines and neural networks using Principal Component Analysis. *Simul. Model. Pract. Theory* **2008**, *16*, 1488–1502. [CrossRef]
118. Durao, R.M.; Mendes, M.T.; Pereira, M.J. Forecasting O_3 levels in industrial area surroundings up to 24 h in advance, combining classification trees and MLP models. *Atmos. Pollut. Res.* **2016**, *7*, 961–970. [CrossRef]
119. Gu, Y.; Li, B.; Meng, Q. Hybrid interpretable predictive machine learning model for air pollution prediction. *Neurocomputing* **2022**, *468*, 123–136. [CrossRef]
120. Czernecki, B.; Marosz, M.; Jędruszkiewicz, J. Assessment of machine learning algorithms in short-term forecasting of pm10 and pm2.5 concentrations in selected polish agglomerations. *Aerosol Air Qual. Res.* **2021**, *21*, 200586. [CrossRef]
121. Qi, Y.; Li, Q.; Karimian, H.; Liu, D. A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory. *Sci. Total Environ.* **2019**, *664*, 1–10. [CrossRef]
122. Zhu, S.; Qiu, X.; Yin, Y.; Fang, M.; Liu, X.; Zhao, X.; Shi, Y. Two-step-hybrid model based on data preprocessing and intelligent optimization algorithms (CS and GWO) for NO_2 and SO_2 forecasting. *Atmos. Pollut. Res.* **2019**, *10*, 1326–1335. [CrossRef]
123. Özceylan, E. Forecasting CO_2 emission of Turkey: Swarm intelligence approaches. *Int. J. Glob. Warm.* **2016**, *9*, 337–361. [CrossRef]
124. Ng, K.Y.; Ngo, D.T.; Chao, P.C.P.; Horng, R.H.; Shieh, J.M. High-Precision Prediction of NOx, NO2 and C6H6 by Multiple Gas Sensors Using a Novel Cascaded MLP-LSTM Model. In Proceedings of the 2024 IEEE SENSORS, Kobe, Japan, 20–23 October 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–4.
125. Chang, Y.S.; Abimannan, S.; Chiao, H.T.; Lin, C.Y.; Huang, Y.P. An ensemble learning based hybrid model and framework for air pollution forecasting. *Environ. Sci. Pollut. Res.* **2020**, *27*, 38155–38168. [CrossRef]
126. Du, P.; Wang, J.; Yang, W.; Niu, T. A novel hybrid fine particulate matter (PM2.5) forecasting and its further application system: Case studies in China. *J. Forecast.* **2022**, *41*, 64–85. [CrossRef]
127. Kouziokas, G.N. SVM kernel based on particle swarm optimized vector and Bayesian optimized SVM in atmospheric particulate matter forecasting. *Appl. Soft Comput.* **2020**, *93*, 106410. [CrossRef]
128. Harishkumar, K.; Yogesh, K.; Gad, I. Forecasting air pollution particulate matter (PM2.5) using machine learning regression models. *Procedia Comput. Sci.* **2020**, *171*, 2057–2066. [CrossRef]
129. Zeinalnezhad, M.; Chofreh, A.G.; Goni, F.A.; Klemeš, J.J. Air pollution prediction using semi-experimental regression model and Adaptive Neuro-Fuzzy Inference System. *J. Clean. Prod.* **2020**, *261*, 121218. [CrossRef]
130. Lyu, T.; Gan, Y.; Zhang, R.; Wang, S.; Li, D.; Zhuo, Y. Development of a Real-Time NOx Prediction Soft Sensor Algorithm for Power Plants Based on a Hybrid Boost Integration Model. *Energies* **2024**, *17*, 4926. [CrossRef]
131. Li, Z.; Bi, J.; Liu, Y.; Hu, X. Forecasting O_3 and NO_2 concentrations with spatiotemporally continuous coverage in southeastern China using a Machine learning approach. *Environ. Int.* **2025**, *195*, 109249. [CrossRef]
132. Dong, J.; Zhang, Y.; Hu, J. Short-term air quality prediction based on EMD-transformer-BiLSTM. *Sci. Rep.* **2024**, *14*, 20513. [CrossRef]
133. Baklanov, A.; Zhang, Y. Advances in air quality modeling and forecasting. *Glob. Transitions* **2020**, *2*, 261–270. [CrossRef]

134. Liao, Q.; Zhu, M.; Wu, L.; Pan, X.; Tang, X.; Wang, Z. Deep learning for air quality forecasts: A review. *Curr. Pollut. Rep.* **2020**, *6*, 399–409. [CrossRef]
135. Liu, H.; Yin, S.; Chen, C.; Duan, Z. Data multi-scale decomposition strategies for air pollution forecasting: A comprehensive review. *J. Clean. Prod.* **2020**, *277*, 124023. [CrossRef]
136. Liu, H.; Yan, G.; Duan, Z.; Chen, C. Intelligent modeling strategies for forecasting air quality time series: A review. *Appl. Soft Comput.* **2021**, *102*, 106957. [CrossRef]
137. Masood, A.; Ahmad, K. A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. *J. Clean. Prod.* **2021**, *322*, 129072. [CrossRef]
138. Méndez, M.; Merayo, M.G.; Núñez, M. Machine learning algorithms to forecast air quality: A survey. *Artif. Intell. Rev.* **2023**, *56*, 10031–10066. [CrossRef]
139. Kryza, M.; Werner, M.; Denby, B.R.; Mu, Q.; Sawiński, T.; Remut, A. Forecasting PM2.5 Concentrations with uEMEP and EMEP4PL for Poland. In *International Technical Meeting on Air Pollution Modelling and its Application*; Springer: Cham, Switzerland, 2021; pp. 193–197.
140. Vovk, T.; Kryza, M.; Werner, M. Using Random Forest to improve EMEP4PL model estimates of daily PM2.5 in Poland. *Atmos. Environ.* **2024**, *332*, 120615. [CrossRef]
141. Krajny, E.A.; Ośródka, L.; Wojtylak, M.J. Application of Doppler sodar in short-term forecasting of PM 10 concentration in the air in Krakow (Poland). *Atmos. Meas. Tech.* **2024**, *17*, 2451–2464. [CrossRef]
142. Sówka, I.; Chlebowska-Styś, A.; Pachurka, Ł.; Rogula-Kozłowska, W.; Mathews, B. Analysis of particulate matter concentration variability and origin in selected urban areas in Poland. *Sustainability* **2019**, *11*, 5735. [CrossRef]
143. Majewski, G.; Szeląg, B.; Mach, T.; Rogula-Kozłowska, W.; Anioł, E.; Bihałowicz, J.; Dmochowska, A.; Bihałowicz, J.S. Predicting the number of days with visibility in a specific range in Warsaw (Poland) based on meteorological and air quality data. *Front. Environ. Sci.* **2021**, *9*, 623094. [CrossRef]
144. Shi, K.; Wu, L. Forecasting air quality considering the socio-economic development in Xingtai. *Sustain. Cities Soc.* **2020**, *61*, 102337. [CrossRef]
145. Guo, Q.; He, Z.; Wang, Z. Simulating daily PM2.5 concentrations using wavelet analysis and artificial neural network with remote sensing and surface observation data. *Chemosphere* **2023**, *340*, 139886. [CrossRef]
146. Ke, H.; Gong, S.; He, J.; Zhang, L.; Cui, B.; Wang, Y.; Mo, J.; Zhou, Y.; Zhang, H. Development and application of an automated air quality forecasting system based on machine learning. *Sci. Total Environ.* **2022**, *806*, 151204. [CrossRef]
147. Cichowicz, R.; Wielgosiński, G.; Fetter, W. Effect of wind speed on the level of particulate matter PM10 concentration in atmospheric air during winter season in vicinity of large combustion plant. *J. Atmos. Chem.* **2020**, *77*, 35–48. [CrossRef]
148. Qin, S.; Liu, F.; Wang, J.; Sun, B. Analysis and forecasting of the particulate matter (PM) concentration levels over four major cities of China using hybrid models. *Atmos. Environ.* **2014**, *98*, 665–675. [CrossRef]
149. Hosseini, S.M.; Saifoddin, A.; Shirmohammadi, R.; Aslani, A. Forecasting of CO_2 emissions in Iran based on time series and regression analysis. *Energy Rep.* **2019**, *5*, 619–631. [CrossRef]
150. Lee, P.; McQueen, J.; Stajner, I.; Huang, J.; Pan, L.; Tong, D.; Kim, H.; Tang, Y.; Kondragunta, S.; Ruminski, M.; et al. NAQFC developmental forecast guidance for fine particulate matter (PM 2.5). *Weather. Forecast.* **2017**, *32*, 343–360. [CrossRef]
151. Sanchez, M.; Pascual, D.; Ramos, C.; Perez, I. Forecasting particulate pollutant concentrations in a city from meteorological variables and regional weather patterns. *Atmos. Environ. Part A Gen. Top.* **1990**, *24*, 1509–1519. [CrossRef]
152. Safdarnejad, S.M.; Tuttle, J.F.; Powell, K.M. Dynamic modeling and optimization of a coal-fired utility boiler to forecast and minimize NOx and CO emissions simultaneously. *Comput. Chem. Eng.* **2019**, *124*, 62–79. [CrossRef]
153. Pagowski, M.; Grell, G.; McKeen, S.; Peckham, S.; Devenyi, D. Three-dimensional variational data assimilation of ozone and fine particulate matter observations: Some results using the Weather Research and Forecasting—Chemistry model and Grid-point Statistical Interpolation. *Q. J. R. Meteorol. Soc.* **2010**, *136*, 2013–2024. [CrossRef]
154. Ofosu-Adarkwa, J.; Xie, N.; Javed, S.A. Forecasting CO_2 emissions of China's cement industry using a hybrid Verhulst-GM (1, N) model and emissions' technical conversion. *Renew. Sustain. Energy Rev.* **2020**, *130*, 109945. [CrossRef]
155. Won, W.S.; Oh, R.; Lee, W.; Ku, S.; Su, P.C.; Yoon, Y.J. Hygroscopic properties of particulate matter and effects of their interactions with weather on visibility. *Sci. Rep.* **2021**, *11*, 16401. [CrossRef]
156. Yang, H.; Zhu, Z.; Li, C.; Li, R. A novel combined forecasting system for air pollutants concentration based on fuzzy theory and optimization of aggregation weight. *Appl. Soft Comput.* **2020**, *87*, 105972. [CrossRef]
157. García, M.V.; Aznarte, J.L. Shapley additive explanations for NO_2 forecasting. *Ecol. Inform.* **2020**, *56*, 101039. [CrossRef]
158. Cinar, Y.G.; Mirisaee, H.; Goswami, P.; Gaussier, E.; Aït-Bachir, A.; Strijov, V. Position-based content attention for time series forecasting with sequence-to-sequence RNNs. In Proceedings of the Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, 14–18 November 2017; Proceedings, Part V 24; Springer: Cham, Switzerland, 2017; pp. 533–544.
159. Vasseur, S.P.; Aznarte, J.L. Comparing quantile regression methods for probabilistic forecasting of NO_2 pollution levels. *Sci. Rep.* **2021**, *11*, 11592. [CrossRef]

160. Martínez-Silva, I.; Roca-Pardiñas, J.; Ordóñez, C. Forecasting SO_2 pollution incidents by means of quantile curves based on additive models. *Environmetrics* **2016**, *27*, 147–157. [CrossRef]
161. Skrobacki, Z.; Dabek, L.E.; Lagowski, P. Modelling and Forecasting of SO_2 Concentration in Atmospheric Air–A Case Study of the City of Krakow. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2019; Volume 471, p. 102057.
162. Sicard, P.; Crippa, P.; De Marco, A.; Castruccio, S.; Giani, P.; Cuesta, J.; Paoletti, E.; Feng, Z.; Anav, A. High spatial resolution WRF-Chem model over Asia: Physics and chemistry evaluation. *Atmos. Environ.* **2021**, *244*, 118004. [CrossRef]
163. Wang, N.; Guo, H.; Jiang, F.; Ling, Z.; Wang, T. Simulation of ozone formation at different elevations in mountainous area of Hong Kong using WRF-CMAQ model. *Sci. Total Environ.* **2015**, *505*, 939–951. [CrossRef]
164. Guariso, G.; Maione, M.; Volta, M. A decision framework for Integrated Assessment Modelling of air quality at regional and local scale. *Environ. Sci. Policy* **2016**, *65*, 3–12. [CrossRef]
165. Grell, G.; Baklanov, A. Integrated modeling for forecasting weather and air quality: A call for fully coupled approaches. *Atmos. Environ.* **2011**, *45*, 6845–6851. Modeling of Air Quality Impacts, Forecasting and Interactions with Climate. [CrossRef]
166. Kim, J.; Jeong, U.; Ahn, M.H.; Kim, J.H.; Park, R.J.; Lee, H.; Song, C.H.; Choi, Y.S.; Lee, K.H.; Yoo, J.M.; et al. New Era of Air Quality Monitoring from Space: Geostationary Environment Monitoring Spectrometer (GEMS). *Bull. Am. Meteorol. Soc.* **2020**, *101*, E1–E22. [CrossRef]
167. Zoogman, P.; Liu, X.; Suleiman, R.; Pennington, W.; Flittner, D.; Al-Saadi, J.; Hilton, B.; Nicks, D.; Newchurch, M.; Carr, J.; et al. Tropospheric emissions: Monitoring of pollution (TEMPO). *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *186*, 17–39. [CrossRef]
168. Li, C.; Krotkov, N.A.; Carn, S.; Zhang, Y.; Spurr, R.J.D.; Joiner, J. New-generation NASA Aura Ozone Monitoring Instrument (OMI) volcanic SO_2 dataset: Algorithm description, initial results, and continuation with the Suomi-NPP Ozone Mapping and Profiler Suite (OMPS). *Atmos. Meas. Tech.* **2017**, *10*, 445–458. [CrossRef]
169. Zeng, J.; Vollmer, B.; Ostrenga, D.; Gerasimov, I. Air Quality Satellite Monitoring by TROPOMI on Sentinel-5P. 2019. Available online: https://ntrs.nasa.gov/api/citations/20180008621/downloads/20180008621.pdf (accessed on 13 August 2025).
170. Jackson, J.M.; Liu, H.; Laszlo, I.; Kondragunta, S.; Remer, L.A.; Huang, J.; Huang, H.C. Suomi-NPP VIIRS aerosol algorithms and data products. *J. Geophys. Res. Atmos.* **2013**, *118*, 12673–12689. [CrossRef]
171. Kellogg, K.; Hoffman, P.; Standley, S.; Shaffer, S.; Rosen, P.; Edelstein, W.; Dunn, C.; Baker, C.; Barela, P.; Shen, Y.; et al. NASA-ISRO Synthetic Aperture Radar (NISAR) Mission. In Proceedings of the 2020 IEEE Aerospace Conference, Big Sky, MT, USA, 7–14 March 2020; pp. 1–21. [CrossRef]
172. Mak, H.W.L.; Laughner, J.L.; Fung, J.C.H.; Zhu, Q.; Cohen, R.C. Improved Satellite Retrieval of Tropospheric NO_2 Column Density via Updating of Air Mass Factor (AMF): Case Study of Southern China. *Remote Sens.* **2018**, *10*, 789. [CrossRef]
173. Benavente, N.R.; Vara-Vela, A.L.; Nascimento, J.P.; Acuna, J.R.; Damascena, A.S.; de Fatima Andrade, M.; Yamasoe, M.A. Air quality simulation with WRF-Chem over southeastern Brazil, part I: Model description and evaluation using ground-based and satellite data. *Urban Clim.* **2023**, *52*, 101703. [CrossRef]
174. Acdan, J.J.M.; Pierce, R.B.; Kuang, S.; McKinney, T.; Stevenson, D.; Newchurch, M.J.; Pfister, G.; Ma, S.; Tong, D. Evaluation of WRF-Chem air quality forecasts during the AEROMMA and STAQS 2023 field campaigns. *J. Air Waste Manag. Assoc.* **2024**, *74*, 783–803. [CrossRef] [PubMed]
175. Christopoulos, J.; Saide, P.; Mohanty, M.; Kim, J.; Judd, L.M.; St Clair, J.M.; Delaria, E.; Wolfe, G.M.; Sebol, A.; Franchin, A.; et al. Model, Satellite, and Airborne NO_2 Comparisons across Southeast Asia: A case study during ASIA-AQ. In Proceedings of the AGU Fall Meeting Abstracts, Washington, DC, USA, 9–13 December 2024; Volume 2024, p. A23L-03.
176. Hsu, Y.C.; Cross, J.; Dille, P.; Tasota, M.; Dias, B.; Sargent, R.; Huang, T.H.K.; Nourbakhsh, I. Smell Pittsburgh: Engaging Community Citizen Science for Air Quality. *ACM Trans. Interact. Intell. Syst.* **2020**, *10*, 1–49. [CrossRef]
177. Jaskulska, A.; Skorupska, K.; Bubrowska, Z.; Kwiatkowska, K.; Stawski, W.; Krzywicki, M.; Kornacka, M.; Kopeć, W. Participatory Action for Citizens' Engagement to Develop a Pro-environmental Research Application. In *Digital Interaction and Machine Intelligence*; Biele, C., Kacprzyk, J., Kopeć, W., Owsiński, J.W., Romanowski, A., Sikorski, M., Eds.; Springer: Cham, Switzerland, 2022; pp. 198–207.
178. Mak, H.W.L.; Lam, Y.F. Comparative assessments and insights of data openness of 50 smart cities in air quality aspects. *Sustain. Cities Soc.* **2021**, *69*, 102868. [CrossRef]
179. Kotsev, A.; Schade, S.; Craglia, M.; Gerboles, M.; Spinelle, L.; Signorini, M. Next Generation Air Quality Platform: Openness and Interoperability for the Internet of Things. *Sensors* **2016**, *16*, 403. [CrossRef]
180. Shahbaz, M.; Gao, C.; Zhai, L.; Shahzad, F.; Khan, I. Environmental air pollution management system: Predicting user adoption behavior of big data analytics. *Technol. Soc.* **2021**, *64*, 101473. [CrossRef]

A.2 Artykuł A2 – Załącznik 2

Urban Climate 59 (2025) 102262

Contents lists available at ScienceDirect

Urban Climate

journal homepage: www.elsevier.com/locate/uclim

Forecasting the air pollution concentration with neural networks

Jarosław Bernacki

Department of Intelligent Computer Systems, Częstochowa University of Technology, al. Armii Krajowej 36, 42-200 Częstochowa, Poland

ARTICLE INFO

Keywords:
Air pollution
Forecasting
Predicting
Machine learning
Deep models
Convolutional neural networks

ABSTRACT

Air pollution is a global problem, which has a major impact on human health. Every year concentrations of many air pollutants cause a large number of deaths. In Europe, particularly Poland, there is poor air quality. In this paper, we deal with forecasting the concentration of air pollutants. We propose four deep learning-based methods for forecasting, which include temporal convolutional network (TCN), Kolmogorov-Arnold Network (KAN), fully convolutional network (FCN), and gated recurrent unit (GRU). Each of the methods is used in three different configurations. We generate predictions for eight air pollutants, from eight cities during a heating season in Poland. Extensive twofold experimental evaluation, combining statistical hypotheses verification and error measures (MAE, MAPE, RMSE) for more than 740 forecast models confirmed high prediction accuracy. Moreover, experiments revealed the advantage of the proposed methods over several state-of-the-art algorithms from the literature.

1. Introduction

Air pollution is a global problem that has a major impact on human health. The concentrations of many air pollutants, especially PM_x are responsible for a large number of deaths each year. In Europe, the air quality is poor in many areas, especially in urban ones Połednik (2022). In many countries, due to the low awareness and understanding of the problem, the negative impact of poor air quality is often ignored. One of the reasons may be insufficient knowledge about this important problem. Industrialization and automobile transportation with fuel combustion contributes to the increase of many air pollutants such as benzene (C_6H_6), carbon monoxide (II) (CO), nitrogen oxide (II) and (IV) (NO, NO_2), nitrogen oxygen (NO_x), ozone (O_3), particulate matter (PM_{10}), sulfur dioxide (SO_2) etc. Therefore, air pollution is monitored and analyzed at air quality monitoring stations. The air quality is evaluated on the basis of the data from air quality monitoring stations.

In Poland, there is in general poor air quality. World Health Organization (WHO) reports that among 50 cities in the European Union with the worst air quality, 33 are located in Poland. Although the reduction in emissions of particulate matter is observed, high concentrations of $PM_{2.5}$ and PM_{10} are still reported, especially in the heating season. However, the air quality is diverse in different Polish cities sow. The most polluted regions in Poland cover the area of the south part of the country, especially the region of Upper Silesia. The cleanest air is observed in the northern part of the country, close to the sea and in Lubusz Land. The criteria and requirements for both air quality and air monitoring are specified by the Chief Inspectorate of Environmental Protection (GIOŚ).

Therefore, monitoring air quality is a very important aspect. One of the problems is the completeness of the monitoring data. Very often the air quality monitoring stations experience some technical difficulties which are revealed by the lack of measurements. Thus, one of the ways is to approximate missing values. This is connected with the task of forecasting. Many papers in the literature focus on forecasting the concentration of air pollutants Liu et al. (2021, 2020); Yang et al. (2020). Used methods include large number of

E-mail address: jaroslaw.bernacki@pcz.pl.

https://doi.org/10.1016/j.uclim.2024.102262
Received 3 September 2024; Received in revised form 12 November 2024; Accepted 21 December 2024
Available online 29 December 2024
2212-0955/

approaches: machine learning Dobrea et al. (2020); Sharma et al. (2021), deep learning Abirami and Chitra (2021); Heydari et al. (2022) or hybrid approach Chang et al. (2020); Mao et al. (2021). However, sometimes the estimation results may not be satisfactory.

Forecasting air pollution is essential for several key reasons. The most important is protecting public health. Air pollutants such as particulate matter, nitrogen oxides, ozone, and sulfur dioxide have a direct impact on human health. They can cause or exacerbate respiratory, cardiovascular, and other diseases. Air quality forecasts allow for taking precautions, such as limiting time spent outdoors, especially for the most vulnerable, such as children, the elderly, or those with chronic diseases. Regularly sharing air pollution forecasts increases public awareness of air quality and its impact on health. This allows people to make informed decisions, such as avoiding outdoor activities on days with high levels of pollution. Air pollution forecasting may also be helpful for crisis management because, in emergency situations, such as forest fires or industrial accidents, air pollution forecasts can help assess the risk and co-ordinate rescue and protection actions for the population.

In this paper, we propose methods for forecasting the concentration of the selected air pollutants in Poland. The analyzed air pollutants include: C_6H_6 (benzene), CO (carbon monoxide (II)), NO (nitrogen oxide (II)), NO_2 (nitrogen oxide (IV)), NO_x (nitrogen oxygen), O_3 (ozone), PM_{10} (particulate matter), SO_2 (sulfur dioxide). These pollutants were analyzed in eight air quality monitoring stations in Poland, located in different Polish cities: Jelenia Góra, Wrocław, Gorzów Wielkopolski, Sulęcin, Zakopane, Gdańsk, Częstochowa, and Koszalin. We have proposed methods utilizing novel approaches for forecasting: temporal convolutional network (TCN), Kolmogorov-Arnold Network (KAN), fully convolutional network (FCN), and gated recurrent unit (GRU). Our goal is to obtain forecasts of the concentration of the mentioned air pollutants. We have realized this, by learning each network with real data collected from GIOŚ, in order to generate predictions of these concentrations. Then, we have compared the results of forecasting with these data, to evaluate the quality of the predictions. Also, each of the proposed networks was run in three different input parameter combinations. As a result, we have obtained more than 740 prediction models, which we have further analyzed. These results were evaluated twofold: by statistical hypotheses verification, and by analyzing the error measures, such as: mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE). The statistical hypotheses verification clearly confirmed that there is no significant difference between the original data and forecasts generated by the proposed methods.

We have also implemented well-known state-of-the-art algorithms for forecasting, such as exponential smoothing, Holt-Winters exponential smoothing, autoregressive integrated moving average (ARIMA), and random forest. Experimental evaluation confirmed that the proposed methods generate more accurate forecasts. The statistical hypotheses verification revealed that in some cases state-of-the-art algorithms generated forecasts with significant statistical difference, compared with the original data. The analysis of MAE, MAPE, and RMSE also indicated that the proposed methods obtained lower values of errors, compared with algorithms from the literature. This leads to the result that the proposed methods generate more accurate forecasts. Such analysis may be useful to estimate the data in case it would be unavailable, for example, due to the technical issues in the air monitoring station.

1.1. *Contribution*

The contribution of this paper is the following. We propose four deep learning-based methods for forecasting air pollutants. Proposed methods include neural networks based on temporal convolutional networks (TCN), Kolmogorov-Arnold Networks (KAN), fully convolutional networks (FCN), and gated recurrent units (GRU). We benchmark proposed methods on the real data collected from GIOŚ of eight selected cities in Poland. Obtained results are analyzed twofold: by statistical hypotheses verification, and by analyzing the error measures. Findings clearly confirm that the proposed methods generate accurate forecasts of the concentration of the air pollutants. The proposed methods' key element that distinguishes them is achieving a lower forecast error compared to the methods described in the literature. In particular, our methods, thanks to modern and recent neural networks (especially relatively new Kolmogorov-Arnold networks), allow for achieving higher forecast accuracy, which we have shown in the comparative analysis.

1.2. *Organization of the paper*

The paper is organized as follows. The next section describes previous and related work. Section 3 presents proposed deep learning methods for forecasting. In section 4 the experimental evaluation of the proposed methods is performed. Finally, the last section concludes this work. At the end of the paper, there is an Appendix that stores the large number of evaluation tables.

2. Previous and related work

A comprehensive survey of different techniques using Artificial Intelligence (AI) for air pollution forecasting is presented in Subramaniam et al. (2022). The authors discuss various methods, including traditional statistical models, like ARIMA/regression models Zhang et al. (2018); Masood and Ahmad (2021); Shams et al. (2021); Zeinalnezhad et al. (2020), artificial neural networks Ahmadi et al. (2023); Cho and Moon (2022); Mishra and Goyal (2015); Shams et al. (2021); Sohn et al. (1999); Usmani et al. (2021), support vector machine Murillo-Escobar et al. (2019); Nilashi et al. (2020); Wang et al. (2017), fuzzy logic Masood and Ahmad (2021), deep learning models Bekkar et al. (2021); Du et al. (2019); Gilik et al. (2022); Heuvelmans et al. (2021); Kurnaz and Demir (2022); Lee et al. (2017); Pak et al. (2020); Saxena (2021), and hybrid models Gu et al. (2022); Qi et al. (2019); Zhu et al. (2018). Considered models are used mainly for forecasting various major pollutants (e.g. $PM_{2.5}$, PM_{10}, O_3, CO, SO_2, NO_2, CO_2). The performance evaluation error indexes include measures R-squared (the coefficient of determination, R^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). A review of hybrid methods useful for air quality forecasting is also presented in Baklanov and Zhang (2020); Liao et al. (2020); Liu et al. (2021); Méndez et al. (2023).

In Hong Zhang et al. (2017) the concentrations of PM_{10} in Taiyuan (China) were analyzed by using the wavelet model of ARMA/ARIMA. The results were modeled with the relative error (RE), root mean square error (RMSE), and normalized root mean square error (NRMSE). According to the results, the values of listed errors are minimal. In Harishkumar et al. (2020) a method for predicting the values of $PM_{2.5}$ using a regression model is proposed. Collected data include real observations from Pingtung station and Newport City in Taiwan. The proposed method utilized the random forest, gradient boosting regression, decision tree regression, and multilayer perceptron regression. The evaluation performance is conducted with MEA and RMSE measures. In Chang et al. (2020) an aggregated LSTM model for air pollution forecasting is considered. Experiments are performed against three popular machine learning methods. For evaluation, the MAE, RMSE, and MAPE measures are used. The research is conducted based on data collected from several places in Taiwan. In Lee et al. (2020) a gradient-boosting-based machine learning approach for predicting the $PM_{2.5}$ concentration in Taiwan is proposed. Obtained results are also compared with Taipei, and London. The findings revealed that Taipei and London have similar prediction results because of the similar city topography. Results were evaluated with the measures R^2 (coefficient of determination), and root mean square error (RMSE).

In Bhatti et al. (2021) a method for predicting the concentration of $PM_{2.5}$ using seasonal autoregressive integrated moving average (SARIMA) is depicted. The evaluation is performed by using the statistical analysis. In Song and Fu (2020) a radial basis neural network is discussed for the air quality prediction in Zhengzhou and Shanghai China. In Shi and Wu (2020), a method for the prediction of a comprehensive air quality index in Xingtai (China) is proposed. Forecasting is realized by utilizing the speed of socio-economic development. Results indicated that the grey multivariable model has high forecasting performance. In Ke et al. (2022) a method for daily predicting $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3, and CO based on data from Beijing, Shanghai, Guangzhou, Chengdu, Xi'an, Wuhan, and Changchun (China) is described. Five machine learning models are used for generating the next 3-days pollution level forecasts. The evaluation confirmed satisfactory results of the predictions.

In sow the analysis of particulate matter concentration in selected Polish cities is discussed. The studies were conducted in the years 2014–2016 in Wrocław and Poznań, and the results revealed that the highest concentration of $PM_{1.0}$, $PM_{2.5}$, and PM_{10} were observed in 2016 in both cities. In Czernecki et al. (2021) a machine learning methods for short-term forecasting of $PM_{2.5}$ and PM_{10} in the biggest Polish cities are discussed. The analysis includes using AIC-based stepwise regression, two tree-based algorithms (random forests and XGBoost), and neural networks. Experimental evaluation confirmed the high accuracy of forecasting by using mentioned methods. In Majewski et al. (2021) a method for predicting the number of days with atmospheric visibility in Warsaw (Poland) based on meteorological and air quality data is proposed. The approach utilizes an Iman–Conover method that can be used to simulate a time series of meteorological and air quality parameters. Experiments enhanced by statistical verification showed that the considered model can effectively determine the number of days in a month with visibility in a specific range. In Kujawska et al. (2022) a number of machine learning methods for predicting the concentration of PM_{10} in Lublin (Poland) is discussed. The following methods were used: linear regression (LR), k-nearest neighbors regression (k-NNR), support vector machine (SVM), regression trees (RT), Gaussian process regression models (GPR), artificial neural network (ANN) and long short-term memory network (LSTM). Experimental evaluation by analyzing the coefficient of determination (R^2) indicated that the artificial neural network achieved the best prediction results over the other methods. Another work with air quality analysis may be found in Gaj et al. (2020); Kryza et al. (2020).

3. Proposed methods

In this section, we describe the architectures of the four networks that might be used for the task of forecasting. Discussed networks include temporal convolutional network (TCN), Kolmogorov-Arnold Network (KAN), fully convolutional network (FCN), and gated recurrent unit (GRU). Before intriducing the proposed methods, let us briefly describe the basics of neural networks when applied to forecasting tasks.

Neural network forecasting relies on the ability of these models to detect complex patterns in data, allowing them to predict future values based on historical observations. Neural networks learn from large data sets, adapting their parameters to minimize forecast errors and increase the accuracy of results. This approach is especially effective in complex problems such as financial forecasting, weather forecasting, or analyzing customer behavior. The process of training a neural network starts with reading input data, which is then processed by subsequent layers of the model, such as convolutional layers, which extract important features from the data. This information is then passed through activation and fully connected layers, which integrate and interpret the detected patterns. Finally, the network generates an output by comparing it with the actual output value. The difference between the generated output and actual output value is called the error, and is minimized by adjusting the weights and parameters of the network to increase the accuracy of the prediction.

Let us briefly summarize the key concepts related to neural network operations. A convolutional layer is responsible for automatically detecting local patterns in data, making it essential for tasks such as image analysis and time series processing. A filter (or weight matrix) is applied in the convolutional layer and moves across input data (e.g., images or time series), identifying patterns and features like edges or textures. The dilation rate is a parameter that extends the filter's range in the convolutional layer, enabling the network to capture broader patterns without increasing the number of parameters. The kernel size, which defines the filter's dimensions, determines the area the filter processes at each step, impacting the resolution of detected details. Padding is the addition of extra values (usually zeros) to the beginning or end of a sequence, allowing convolutional filters to analyze the entire length of the series, including its beginnings and ends. This allows the network to preserve the full length of the original sequence and better detect patterns at its ends. A fully connected layer has connections of each neuron to every neuron from the previous layer, integrating information from different processing stages. An activation function is a nonlinear function in neurons that decides how signals are transmitted to the next layer, allowing the network to model complex data relationships. One of the most common activation functions is the Rectified Linear Unit (ReLU). It returns the input value if it is positive, or zero if it is negative, which helps the neural network converge faster. Another activation function called the Scaled Exponential Linear Unit (SELU), normalizes activation values within the network, stabilizing training. Dropout is a regularization technique that randomly deactivates neurons during training to prevent overfitting. Finally, the learning rate, which defines the size of weight adjustments in each optimization step, is crucial for achieving a stable and efficient training process.

3.1. Temporal convolutional network

Temporal Convolutional Networks (TCNs) are a type of convolutional neural network (CNN) designed to process sequential data such as time series. Unlike recurrent neural networks (RNNs), TCNs use convolutional operations with dilation, which allows them to efficiently model long-term temporal dependencies without the problems associated with decaying gradient Bai et al. (2018). The structure of a TCN usually includes multiple convolutional layers using dilated convolutions to capture information from distant time points in the sequence. The proposed TCN consists of six convolutional layers, which can be described as follows:

(1) The first 1D convolution layer with kernel size 3, SELU as activation function, and dropout;
(2) Second 1D convolution layer with dilation rate 2, kernel size 3, SELU as activation function and dropout;
(3) Third 1D convolution layer with dilation rate 4, kernel size 3, SELU as activation function and dropout;
(4) Fourth 1D convolution layer with dilation rate 8, kernel size 3, SELU as activation function and dropout;
(5) Fifth 1D convolution layer with dilation rate 16, kernel size 3, SELU as activation function and dropout;
(6) The last layer reduces the number of channels to the desired output using 1×1 convolution, generating a prediction tensor for the sequence.

TCNs are particularly useful in tasks such as video segmentation Lea et al. (2017, 2016) or time series forecasting, where it is important to capture the dependencies between temporally distant elements. In our approach, the input data is processed using a sliding window of a defined size, which allows for generating forecasts based on the entire time sequence. This approach enables TCNs to be applied to a variety of forecasting tasks, effectively combining low- and high-level temporal features.

We propose to use a non-linear SELU function Klambauer et al. (2017), which is a variant of the ReLU function Nair and Hinton (2010) and is defined as presented as Equ 1.

$$\mathrm{SELU}(x) = \lambda \begin{cases} x & \text{for} \quad x > 0 \\ \alpha exp(x) - \alpha & \text{for} \quad x \leq 0 \end{cases} \tag{1}$$

where $\alpha \approx 1.6733$ and $\lambda \approx 1.0507$. Dropout regularization is employed to prevent overfitting by randomly setting a fraction of the activations to zero during training. Additionally, the dilation rate in a convolutional layer specifies how many input values are skipped between each step of the convolution. A dilation rate of 1 corresponds to a standard convolution, whereas higher rates result in skipping more values between the convolutions. The structure of the proposed network is presented in Fig. 1.

3.2. Kolmogorov-Arnold network

Kolmogorov–Arnold Networks (KAN), named after the Kolmogorov–Arnold theorem, which proves that any continuous function of several variables can be approximated by a superposition of continuous one-dimensional functions, were presented in 2024 Liu et al.

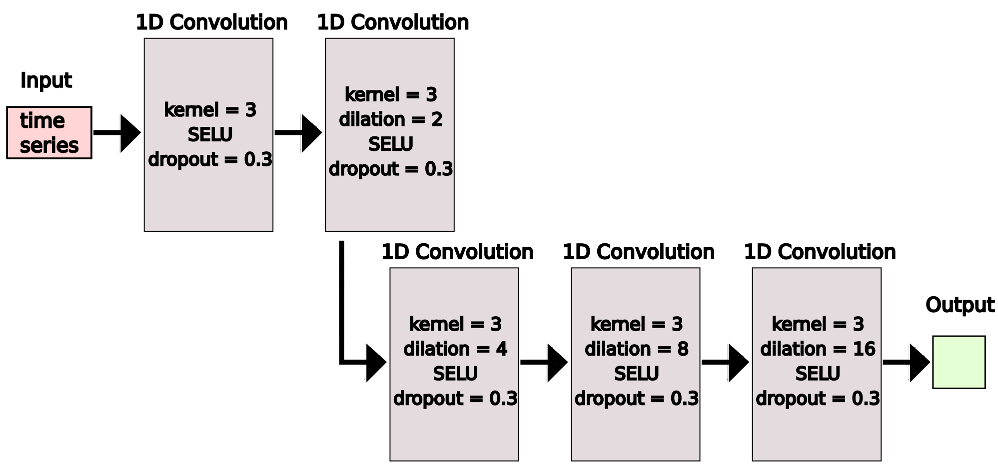


Fig. 1. The structure of the temporal convolutional network (TCN).

(2024). This type of network consists of fully connected layers, where each neuron in one layer is connected to every neuron in the next layer. The KAN architecture allows for capturing more complex patterns in the data compared to traditional neural networks, making it suitable for tasks that require precise modeling of nonlinear relationships, such as modeling complex mathematical functions, financial data analysis, or signal processing.

Each layer performs a linear transformation of the input data, learning a set of weights and biases during training. Furthermore, these layers use a nonlinear activation function, usually sinusoidal, which allows for modeling complex, nonlinear relationships in the data. The sinusoidal function enables the network to reflect complex patterns, according to the Kolmogorov-Arnold theorem theory.

We propose the following structure of the Kolmogorov-Arnold Network:

(1) The first fully connected layer (FC1) with 64 neurons, sinusoidal activation function and dropout;
(2) The second fully connected layer (FC2) with 128 neurons, sinusoidal activation function and dropout;
(3) The third fully connected layer (FC3) with 256 neurons, sinusoidal activation function and dropout;
(4) The fourth fully connected layer (FC4) with 512 neurons, sinusoidal activation function and dropout;
(5) The last layer takes as input the output from FC4 and generates the final prediction.

The network takes a time series as input, processing it with a sliding window of a defined size. The values from subsequent windows are passed through network layers. Finally, the network generates a prediction based on the values from the window. The size of the window can be freely defined. The structure of the proposed network is presented in Fig. 2.

3.3. Fully convolutional network

A fully convolutional network (FCN) is a type of neural network that uses only convolutional layers, eliminating the fully connected layers that are traditionally used in classical CNNs. FCNs were originally designed for image segmentation tasks, where they allow for assigning labels to each pixel of the image. However, their flexible architecture also allows for efficient processing of time series, especially in the context of tasks such as prediction and pattern detection.

We propose the following fully convolutional network structure:

(1) A first convolutional layer with 32 filters, kernel size 3, SELU as activation function, and dropout, transforming the input size into the number of filters;
(2) A second convolutional layer with 64 filters, kernel size 3, SELU as activation function, and dropout;
(3) Third convolutional layer with 128 filters, kernel size 3, SELU as activation function and dropout;
(4) Fourth convolutional layer with 256 filters, kernel size 3, SELU as activation function and dropout;
(5) The last convolutional layer with 512 filters, kernel size 1, SELU as activation function, and dropout, transforming the number of filters into the final result.

Convolutional layers are extremely effective at detecting local patterns in the input data, which is especially important in time series processing, where preserving information about local temporal dependencies is crucial. Using smaller kernel sizes allows the network to capture more detailed patterns in the data while increasing the number of filters in subsequent layers allows the modeling

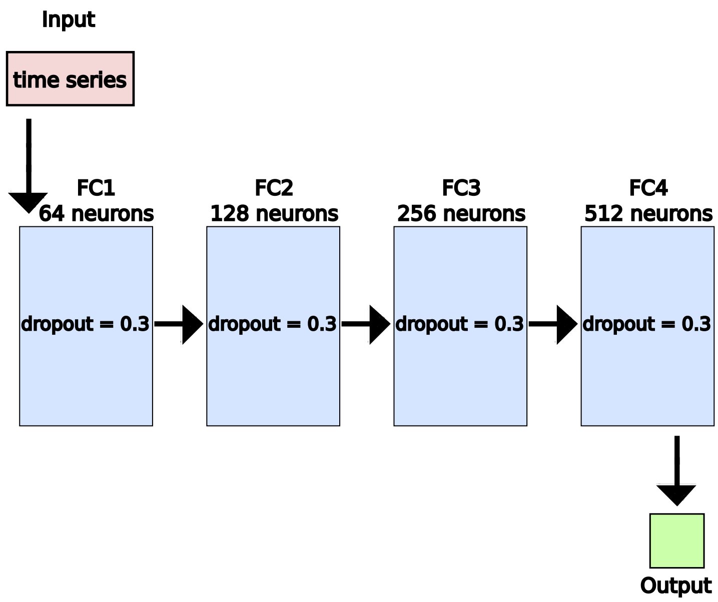


Fig. 2. The structure of the Kolmogorov-Arnold network (KAN).

increasingly complex features. Using convolutional layers instead of fully connected layers for classification offers several advantages. Each neuron in the convolutional layer is connected only to a local region of the previous layer, called the receptive field, which preserves spatial information and eliminates the need to flatten the input data. Convolutional filters slide over the input data, applying the same weights to different positions, which significantly reduces the number of parameters compared to fully connected layers, reducing the risk of overfitting and increasing computational efficiency.

To further reduce the risk of overfitting, we propose to use a dropout regularization technique that randomly excludes a percentage of neurons during training, preventing the model from becoming too dependent on specific neurons. The network takes a time series as input, processing it with a sliding window of a defined size. The values from subsequent windows are passed through network layers that gradually learn more complex patterns in the data. Finally, the network generates a prediction based on the values from the window. The size of the window can be defined arbitrarily. The structure of the proposed network is presented in Fig. 3.

3.4. Gated recurrent unit network

Gated recurrent unit (GRU) is a type of recurrent neural network (RNN) designed by Cho et al. in 2014 Chung et al. (2014). GRU, like Long Short-Term Memory (LSTM), was designed to solve the vanishing gradient problem that made it difficult to train classical RNNs when working with long sequences of data. Unlike LSTM, GRU has a simplified architecture, which makes it less computationally complex and often more efficient.

A GRU network takes input data with parameters: *input size*, *hidden size*, and *output size*. Input size and output size refer to the dimensions of the input and output data, respectively, while hidden size refers to the number of GRU units in a layer.

The structure of the proposed GRU network can be presented as follows:

(1) A layer with 64 GRU units;
(2) A linear layer that maps the last hidden state to the output size.

The key elements of the GRU architecture are two gates: an update gate and a reset gate. These gates control the flow of information through the network, allowing it to adaptively remember or forget information at each time step.

The reset gate determines how much information from the previous hidden state h_{t-1} should be "forgotten" when processing the current time step. Its operation can be described by the following formula:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

where σ is the sigmoid function and W_r is the weight of the reset gate.

The update gate controls to what extent the new state $\widehat{h}_t$ should replace the previous hidden state h_{t-1}, which can be expressed as:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

where W_z is the weight of the update gate.

The final new hidden state h_t is a combination of the previous state and the new information:

$$h_t = z_t \cdot h_{t-1} + (1 - z_t) \cdot \widehat{h}_t$$

The hidden states in the GRU store internal representations of the sequence, which are updated at each time step. This allows the GRU to "remember" information from previous time steps, which is crucial in processing sequences where historical context is

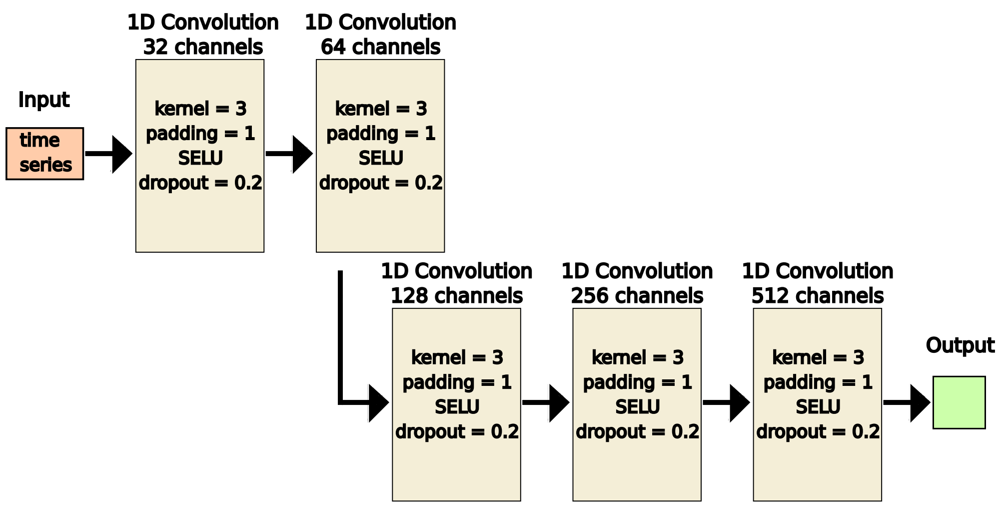


Fig. 3. The structure of the fully convolutional network (FCN).

important. The linear layer at the end of the GRU network transforms the last hidden state into the output prediction. This layer acts as a function that maps a representation of the sequence to a target output, allowing the model to generate predictions based on patterns detected in the input sequence.

In practice, the GRU network takes data sequences as input and processes them using a window of a defined size. Each sequence is passed through GRU which updates the hidden states. Once the entire sequence has been processed, the last hidden state is used to generate the final prediction. The size of the input time window can be adjusted. The structure of the GRU is presented in Fig. 4.

4. Experimental evaluation

In this section, we benchmark the proposed methods by performing forecasting simulations based on the real data from the GIOŚ. The data from the GIOŚ represent values of concentrations $(\mu g/m^3)$ of the particular air pollutants. More precisely, our goal is to learn the proposed methods by the data representing different concentrations of air pollutants from several Polish cities (let us further name it *original data*) and to compare results generated by the proposed methods with the original data. This made it possible to evaluate the forecasting quality of the proposed methods.

4.1. Collected data and area description

The forecasting evaluation was conducted with the following air pollutants: C_6H_6 (benzene), CO (carbon monoxide (II)), NO (nitrogen oxide (II)), NO_2 (nitrogen oxide (IV)), NO_x (nitrogen-oxygen), O_3 (ozone), PM_{10} (particulate matter), SO_2 (sulfur dioxide). These pollutants were analyzed in the following air quality monitoring stations in Poland: DsJelGorOgin (Jelenia Góra), DsWrocWybCon (Wrocław), LuGorzKosGdy (Gorzów Wielkopolski), LuSulecDudka (Sulęcin), MpZakopaSien (Zakopane), PmGdaLeczkow (Gdańsk), SlCzestoBacz (Częstochowa), and ZpKosz (Koszalin). Please note that the data representing ZpKosz comes from the two air quality monitoring stations in Koszalin: ZpKoszArKraj and ZpKoszChopin. This combination was essential because ZpKoszArKraj does not monitor the concentration of O_3 and SO_2.

We use the codes of the air stations and follow them in the further part of the paper; in the round brackets, there are names of the cities, in which a particular air station is located. We have used the aforementioned data because they are characterized by complete accessibility. Moreover, all listed air pollutants are characterized by hugely different concentrations, so it allows for a better evaluation of the proposed methods. The location of the air stations can be seen in Fig. 5.

We present forecasting results from the hourly data collected from 2 January 2022 (heating season in Poland) for all listed air pollutants from 0:00 to 23:00 h. Thus, the input data for each of the proposed methods is equal to 24. The accuracy of forecasting on data from other days is similar, so we skipped their presentation for clarity. The data for benzene (C_6H_6) is unavailable for LuSulecDudka and MpZakopaSien because their air monitoring stations do not analyze this air pollutant. The data are not pre-processed (e.g. normalized) – the proposed methods are learned by the original data directly.

4.2. Usage of the proposed methods, meta parameters, and implementation

The proposed methods were used in the following manner. Each of them was implemented to take as input multiple values and produce the output of one single value, which was a prediction. The number of input multiple values was different – there were conducted a set of experiments in which each network (TCN, KAN, FCN, and GRU) generated predictions based on 3, 5, and 7 input values (let us further name these values window). So, to calculate predictions for input data, the window was sliding over it and the specific method was producing the prediction. Therefore, using 8 air pollutants from 8 air quality monitoring stations of 4 proposed methods (each method run threefold) generated a 744 series of predictions.

The meta parameters for all networks are similar. Each network was learned for 1000 epochs, the Adam optimizer was used. Almost all networks were learned with a learning rate equal to 0.001, only for the KAN we have used 0.01. The dropout was set to 0.3 for TCN and KAN; in the case of FCN we used 0.2, and the GRU did not use the dropout. The meta parameters are summarized in Table 1.

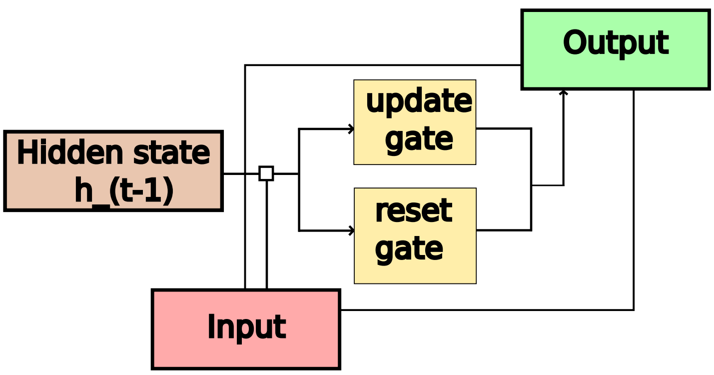


Fig. 4. The structure of the gated recurrent unit (GRU).

Fig. 5. The location of the air quality monitoring stations in Poland used in the experiment.

The meta parameters values were selected experimentally, which is a common approach in complex neural network applications. Iterative testing and analysis of results allowed to find settings that provide adequate model accuracy, lowering the prediction error measures while maintaining computational efficiency. However, due to the complexity of such experiments, their presentation is skipped in the manuscript for clarity.

The scripts for the proposed methods were implemented under Python programming language with PyTorch framework with Nvidia CUDA support. The scripts for statistical verification were implemented in Matlab. Experiments were held on a personal computer equipped with an Intel Core i5–10,400@2.9GHz CPU with 32 gigabytes of RAM and a Nvidia RTX 3060 GPU with 12 gigabytes of video memory.

We compare the proposed methods both with the original data, but also with the forecasts generated by the following state-of-the-art algorithms: exponential smoothing Gardner Jr. (1985), Holt-Winters exponential smoothing Lima et al. (2019), autoregressive integrated moving average (ARIMA) Shumway and Stoffer (2017), and random forest Rigatti (2017). The implementation of state-of-the-art algorithms was realized in Python programming language using built-in libraries: `statsmodels` library which offers a great number of different statistical methods, including exponential smoothing, Holt-Winters exponential smoothing and ARIMA algorithms and the `sklearn` library for the random forest algorithm. Let us summarize the parameters of these algorithms.

(1) Exponential smoothing: this algorithm enables setting the smoothing level which was set to 0.2. This parameter determines the weight of the new observations compared to older observations. No optimization is used.
(2) Holt-Winters exponential smoothing: the algorithm assumes that the data shows seasonality with a seasoning period set to 12 and also assumes that the data shows seasonality. The built-in optimization is enabled.
(3) ARIMA: this algorithm takes as input the three values: (p, d, q). Their usage is in the following manner: $p = 5$, which denotes taking 5 last observations for the prediction; $d = 1$ enables the differentiation only once; $q = 0$ states that the algorithm does not use the moving average (MA) for generating predictions.
(4) Random forest: the algorithm was used with the following parameters. The number of observations used for predictions is 5, the other parameters were set as the following:

Table 1
Meta parameters of the proposed networks.

	TCN	KAN	FCN	GRU
input (# of values)	3/5/7	3/5/7	3/5/7	3/5/7
output (# of values)	1	1	1	1
dropout	0.3	0.3	0.2	not used
learning rate	0.001	0.01	0.001	0.001
optimizer	Adam	Adam	Adam	Adam
# of epochs	1000	1000	1000	1000

- Maximum number of the trees in the forest: 200;
- Maximum depth of the tree: 30;
- Minimal number of samples for splitting the node: 10;
- Minimal number of samples in the leaf: 4.

Technical information about the work and logic of state-of-the-art algorithms' may be found in the aforementioned papers.

4.3. Forecasting examples – visual inspection

Due to the very large number of forecast models (744 figures), we skip presenting them in the paper. We only show two selected forecast models that may be seen in Fig. 6 and 7.

Visual inspection confirms the reliability of the proposed methods. The forecasts of the SO_2 in the DsJelGorOgin air monitoring station generated by the GRU are almost overlapped with the original data (Fig. 6). A similar situation is observed in the case of TCN predictions for CO in ZpKosz (Fig. 7). However, during hours 14–16, some inaccuracies may be observed. What is interesting, less accurate turned out to be the TCN with a window of size 7. Nevertheless, this does not have a negative impact on the overall forecasting accuracy. Detailed numerical analysis of the forecast models we present in the next subsections.

4.4. Statistical verification

The statistical verification relying on hypothesis testing allows for answering the question, of whether the analyzed data differ from each other significantly. We have analyzed, if the original data from GIOŚ differ significantly from the forecasts obtained by the proposed methods. We have also examined if our predictions differ from the forecasts generated by state-of-the-art algorithms. All tests were performed at the $\alpha = 0.05$ significance level.

4.4.1. Normality tests

First of all, the normality tests were performed. We have used the Shapiro-Wilk test to determine if the analyzed data perform the normal distribution. The hypotheses are defined as following Shapiro and Wilk (1965):

- $\mathscr{H}_0$: The analyzed data performs the standard distribution;
- $\mathscr{H}_1$: The analyzed data does not perform the standard distribution.

To accept or reject the null hypothesis, it is essential to calculate the Shapiro-Wilk test statistic S, and then calculate the p-value (let us further name it shortly p). The p is compared with the significance level, and if greater, then there is no reason to reject the null hypothesis $\mathscr{H}_0$. Otherwise, the alternative hypothesis $\mathscr{H}_1$ is accepted. Alternatively, if S is greater than the test critical value, the $\mathscr{H}_0$ might be accepted.

We have checked all the data, if they perform the normal distribution. The critical value for 24 input samples is equal to 0.916. For clarity, we only present results hypotheses values as Table 2. Full results with the values of p and test statistics S are presented in the Appendix as Table A.9-A.16.

The results clearly indicate that the great majority of the analyzed data does not perform the normal distribution. In most cases, the

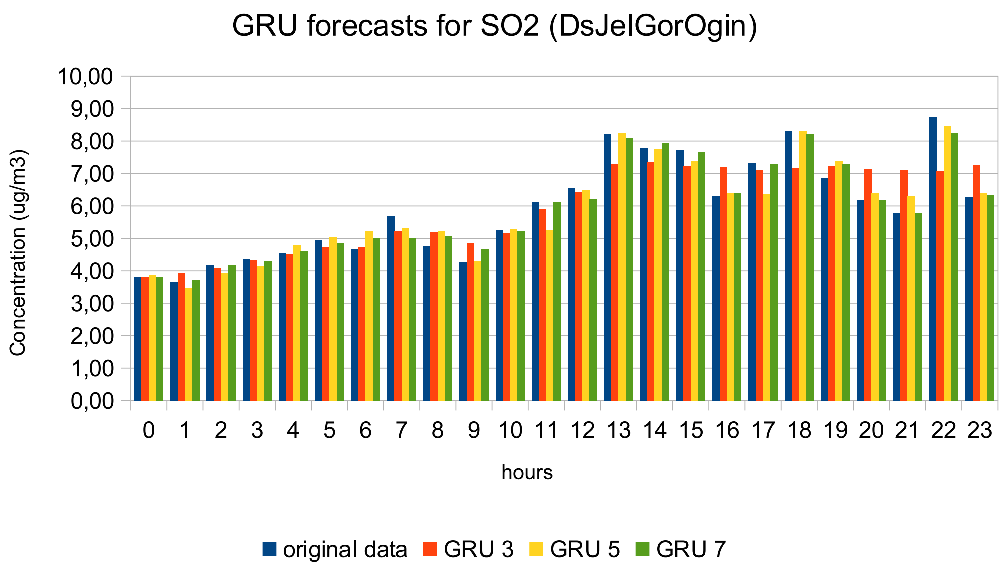


Fig. 6. GRU forecasts compared with the original data for SO_2 (DsJelGorOgin)

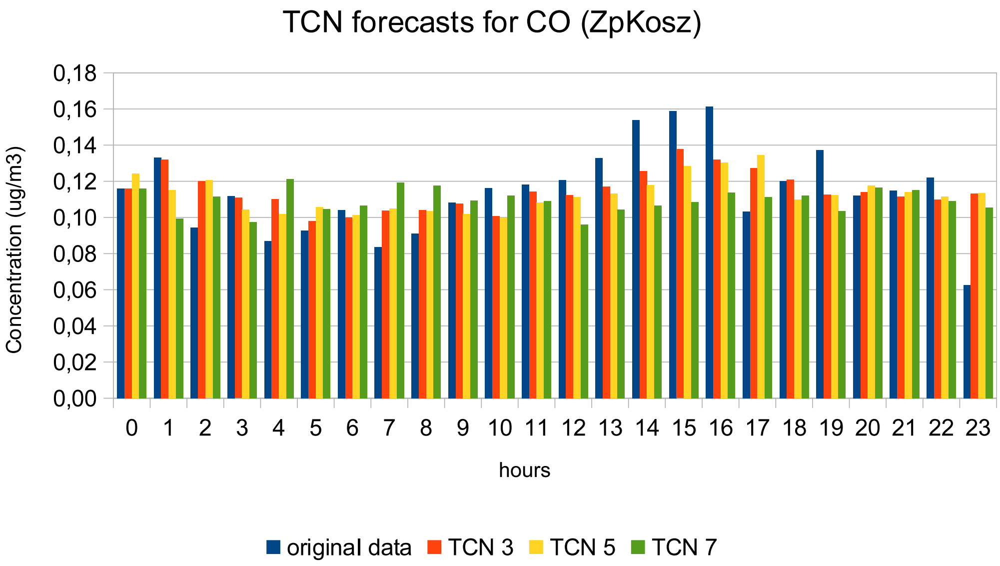


Fig. 7. TCN forecasts compared with the original data for CO (ZpKosz)

p is around 0.00, so the hypothesis $\mathscr{H}_0$ should be rejected. Another confirmation is that the statistical test S is lower than the critical value, so there is no reason to accept the $\mathscr{H}_0$. Therefore, for further analysis, the non-parametric Kruskal-Wallis ANOVA test was used.

4.4.2. Kruskal-Wallis ANOVA

The non-parametric Kruskal-Wallis ANOVA (also called Kruskal-Wallis one-way analysis of variance for ranks) is an extension of the U Mann-Whitney test for more than two populations. This test is used for the verification of the insignificance of differences of the medians in $k > 2$ populations Kruskal and Wallis (1952); Kruskal (1952). The hypotheses for k groups are defined as follows:

- $\mathscr{H}_0$: $\theta_1 = \theta_2 = \cdots = \theta_k$ (all the medians are equal);
- $\mathscr{H}_1$: Not all medians θ_j are equal ($j = 1, 2, \ldots, k$).

To accept or reject the null hypothesis, it is essential to calculate the test statistic F, and then calculate the p. The p is compared with the significance level, and if greater, then there is no reason to reject the null hypothesis $\mathscr{H}_0$. Otherwise, the alternative hypothesis $\mathscr{H}_1$ is accepted. Alternatively, we can compare the test statistic F with the test critical value. If the test statistic is greater than the critical value, the $\mathscr{H}_0$ should be rejected.

Experiment I – Forecasts comparison with original data. In this experiment, we statistically compare the medians of the original data and the forecasts generated by the proposed methods. Since we analyze 13 groups, the number of degrees of freedom between groups is equal to $13 - 1 = 12$. The number of degrees of freedom within groups is equal to 299. Therefore, the critical value is equal to 1.84. To accept the $\mathscr{H}_0$, the value of the test statistic should not exceed the critical value. Results of the Kruskal-Wallis ANOVA test statistic are presented in Table 3.

Results indicate that in no case the values of the test statistics F do not exceed the critical value of the test. What is more, in the majority of cases it is much lower. Therefore, the hypothesis $\mathscr{H}_0$ may be accepted, so it leads to the corollary, that all the forecasts obtained by the proposed methods do not differ significantly from the original data per each air pollutant. Thus we can assume that the forecasts are accurate for all air quality monitoring stations.

Experiment II – Forecasts comparison with original data and state-of-the-art algorithms. In this experiment, we statistically compare the medians of the original data both with the forecasts calculated by the proposed methods and forecasts generated by state-of-the-art algorithms. In this experiment, we analyze 17 groups, and the number of degrees of freedom between groups is equal to $17 - 1 = 16$. The number of degrees of freedom within groups is equal to 391. Thus, the critical value is equal to 1.74. Results of the Kruskal-Wallis ANOVA test statistic are presented in Table 4.

The obtained results suggest that in the majority of the cases, the test statistic F is greater than the critical value (1.74). This means that the hypothesis $\mathscr{H}_0$ should be rejected, and the analyzed data differ significantly. The most demanding was the LuSulecDudka air quality station, where all data turned out to be different from each other. From the air pollutants, the most challenging was SO_2, where in all cases the hypothesis of the median equality was rejected. The very low accuracy of forecasting is noticed for O_3 for DsWrocWybCon, and SlCzestoBacz, where the differences between the analyzed data are very large. The most accurate forecasts are observed in the case of C_6H_6 in DsJelGorOgin, where the value of the F statistic was equal to 0.02. The most stable was the PmGdaLeczkow air station, where only O_3 and SO_2 were different.

To check the specific tendencies, we can perform the post-hoc analysis.

Experiment II – Post-hoc analysis. Due to the fact that Kruskal-Wallis ANOVA only generally indicates, whether the groups differ (or not) significantly without pointing precise differences, the post-hoc analysis may be calculated. The post-hoc indicates the sum of

Table 2

Hypotheses of the normality tests. 0 denotes that data represent normal distribution, and 1 denotes the difference from the normal distribution. The number next to the name of the proposed method denotes the size of the window (e.g. TCN 3 denotes the TCN with a window of size 3; FCN 7 denotes the FCN with a window of size 7); exp. sm. denotes exponential smoothing algorithm; exp. (Holt) denotes Holt-Winters exponential smoothing algorithm; ARIMA stands for autoregressive integrated moving average algorithm; r. forest denotes the random forest algorithm.

	original	TCN 3	TCN 5	TCN 7	KAN 3	KAN 5	KAN 7	FCN 3	FCN 5	FCN 7	GRU 3	GRU 5	GRU 7	exp. sm.	exp. (Holt)	ARIMA	r. forest
DsJelGorOgin																	
C_6H_6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CO	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	1
NO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
NO_2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1
NO_x	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1
O_3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
PM_{10}	0	0	1	0	1	1	1	1	1	1	1	1	1	0	0	0	1
SO_2	0	0	0	0	1	1	1	1	1	1	1	0	0	1	0	0	1
DsWrocWybCon																	
C_6H_6	1	1	1	1	1	1	0	1	0	1	0	1	1	0	0	0	1
CO	1	0	1	0	1	1	1	0	1	0	0	0	0	0	1	1	1
NO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NO_2	0	0	1	1	0	0	0	0	0	0	0	1	0	1	0	0	1
NO_x	0	1	1	0	0	0	0	0	1	0	0	1	1	1	0	0	1
O_3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PM_{10}	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	1
SO_2	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	1	0
LuGorzKosGdy																	
C_6H_6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
CO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NO	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1
NO_2	0	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0
NO_x	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1
O_3	0	0	0	0	1	1	1	0	1	0	1	1	1	0	0	0	0
PM_{10}	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
SO_2	0	1	0	0	1	1	1	1	1	1	1	1	1	0	0	0	1
LuSulecDudka																	
CO	0	0	0	0	1	1	1	0	0	1	1	0	1	1	0	0	1
NO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NO_2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NO_x	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
O_3	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1
PM_{10}	0	1	0	0	1	0	1	1	1	1	1	1	1	0	0	0	0
SO_2	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1
MpZakopaSien																	
CO	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1
NO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NO_2	0	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	1
NO_x	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
O_3	1	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	1
PM_{10}	1	1	1	1	0	1	0	1	1	1	1	1	1	0	1	1	1
SO_2	1	1	0	0	1	0	1	1	0	1	1	0	1	0	0	0	1
PmGdaLeczkow																	
C_6H_6	1	0	1	1	0	1	0	1	1	1	1	1	1	0	0	1	0
CO	0	0	0	0	1	1	1	0	0	0	0	0	1	1	0	1	0
NO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1

(*continued on next page*)

Table 2 (*continued*)

	original	TCN 3	TCN 5	TCN 7	KAN 3	KAN 5	KAN 7	FCN 3	FCN 5	FCN 7	GRU 3	GRU 5	GRU 7	exp. sm.	exp. (Holt)	ARIMA	r. forest
NO_2	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
NO_x	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
O_3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PM_{10}	0	0	0	0	1	1	0	1	1	1	1	1	1	0	0	0	1
SO_2	0	0	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1
SlCzestoBacz																	
C_6H_6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NO_2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NO_x	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
O_3	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
PM_{10}	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1
SO_2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
ZpKosz																	
C_6H_6	0	1	0	0	0	1	0	0	0	0	0	0	0	1	1	1	1
CO	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1
NO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NO_2	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
NO_x	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
O_3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PM_{10}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
SO_2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 3
Values of test statistic F of the Kruskal-Wallis ANOVA test – forecasts comparison with original data.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2
DsJelGorOgin	0.01	0.23	0.01	0.12	0.15	0.34	0.88	0.33
DsWrocWybCon	0.14	0.04	0.11	0.15	0.21	0.52	0.61	0.37
LuGorzKosGdy	0.03	0.02	0.06	0.65	0.85	0.30	0.02	0.71
LuSulecDudka	n/a	0.29	0.17	0.27	0.21	0.88	0.38	0.29
MpZakopaSien	n/a	0.04	0.18	1.00	1.46	0.08	0.33	0.03
PmGdaLeczkow	0.06	1.24	0.03	0.05	0.05	1.23	0.10	0.89
SlCzestoBacz	0.06	0.26	0.12	0.66	0.02	0.65	0.28	0.08
ZpKosz	0.18	0.52	0.03	0.06	0.03	0.55	0.06	0.10

Table 4
Values of test statistic F of the Kruskal-Wallis ANOVA test – forecasts comparison with original data and state-of-the-art algorithms.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2
DsJelGorOgin	0.02	0.22	2.33	3.36	6.11	0.82	2.96	2.88
DsWrocWybCon	0.14	0.06	1.16	2.99	1.15	14.30	7.94	2.93
LuGorzKosGdy	0.24	0.38	3.08	8.86	1.95	5.15	3.89	3.92
LuSulecDudka	n/a	21.50	5.51	2.97	6.96	7.33	80.29	12.62
MpZakopaSien	n/a	3.39	1.98	3.09	4.76	0.74	0.66	4.19
PmGdaLeczkow	0.14	0.85	0.04	1.73	1.30	5.84	0.13	7.40
SlCzestoBacz	4.80	0.24	0.11	0.75	5.19	16.50	5.37	3.96
ZpKosz	5.94	0.51	0.83	0.98	2.87	3.92	5.94	2.07

the ranks of the particular groups, so one may deduce the precise tendencies between them. In Table 5 the values of the sum of ranks for the particular data are described.

Results point out that the sum of ranks between the original input data and the forecasts generated by the proposed methods are similar (as it results from Experiment I). The main differences may be observed in the case of state-of-the-art algorithms. For example, ranks for DsJelGorOgin, DsWrocWybCon, LuGorzKosGdy, LuSulecDudka, and SlCzestoBacz are clearly higher than in the case of the original and the proposed methods. On the other hand, ranks for PmGdaLeczkow, and ZpKosz are lower. This may be interpreted that higher rank values indicate that obtained forecast values are bigger. Similarly, lower rank values point to PmGdaLeczkow, and ZpKosz mean that their forecasted values are smaller than the original and proposed methods. This analysis confirms that state-of-the-art algorithms do not produce accurate forecasts, compared with the proposed methods. To conclude, the overall number of forecast models generated by the proposed methods was equal to 762; the number of forecast models by state-of-the-art algorithms was 248. None of the forecasts generated by proposed networks was significantly different from the original data; 61 forecast models (25 %) by algorithms from the literature were significantly different from the original data. This is another confirmation of the accuracy of the proposed methods.

4.5. *Results of error measures*

We compare the forecasts obtained by the proposed and state-of-the-art methods by calculating estimators including mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE). The MAE is defined as shown in eq. 2:

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} \left| \frac{y_t - x_t}{n} \right| \tag{2}$$

where x_t is the actual value, y_t is a predicted value and n is the number of observations.

The MAPE may be calculated as shown in eq. 3:

Table 5
Post-hoc analysis of sum of ranks for Experiment II.

	original	proposed	state-of-the-art
DsJelGorOgin	187	188	260
DsWrocWybCon	194	187	263
LuGorzKosGdy	200	198	227
LuSulecDudka	188	194	238
MpZakopaSien	207	206	201
PmGdaLeczkow	231	223	144
SlCzestoBacz	200	192	245
ZpKosz	215	212	181

$$\mathrm{MAPE} = 100\frac{1}{n}\sum_{t=1}^{n}\left|\frac{x_t - y_t}{x_t}\right| \tag{3}$$

where x_t is the actual value, y_t is a predicted value and n is the number of observations.

The RMSE might be determined as shown in eq. 4:

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} \tag{4}$$

where:

$$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n}(x_t - y_t)^2 \tag{5}$$

x_t is the actual value, y_t is a predicted value and n is the number of observations.

Due to the large number of obtained results, we present tables with MAE, MAPE, and RMSE values in the Appendix. Full results with detailed MAE values per each air pollutant of every air station are presented as Table A.17-A.24; results of MAPE are presented in Table A.25-A.32, and results of RMSE we describe in Table A.33-A.40.

4.5.1. MAE analysis

In Table 6, the mean values of MAE may be observed (based on all air pollutants in each air quality station), both for all the proposed methods, and state-of-the-art algorithms.

As may be seen, the proposed methods obtain lower MAE values than state-of-the-art algorithms in all cases. In nearly all cases MAE is less than 0.1, however, in the case of MpZakopaSien and SlCzestoBacz, obtained values are slightly greater. According to the minimum and maximum values of the concentration of the particular air pollutants, we may find such results satisfactory. State-of-the-art algorithms achieved much higher MAE values which in all cases exceeded 0.1. The greatest values are observed for LuSulecDudka, SlCzestoBacz, and ZpKosz – 0.23, 0.34, and 0.23, respectively. Of course the greater the MAE value, the less forecast accuracy. This confirms that the proposed methods are much more accurate than methods from the literature. Detailed MAE values for each air pollutant with its minimum and maximum value of concentration in the original data for every air quality monitoring station may be seen in the Appendix as Table A.17-A.24.

4.5.2. MAPE analysis

The mean values of MAPE, calculated by each air pollutant per station, are presented in Table 7, both for all the proposed methods, and state-of-the-art algorithms. The lower the MAPE value, the better the algorithm accuracy.

The proposed methods consistently outperform the state-of-the-art methods across all air quality monitoring stations. In general, the MAPE of the proposed methods does not exceed 16.00 (only MpZakopaSien obtained almost 31.00). The state-of-the-art algorithms reached much worse results. The biggest differences compared to the proposed methods are observed in the case of MpZakopaSien and ZpKosz, where the MAPE achieved 74.39 and 72.30, respectively. For all air quality stations, the MAPE exceeded the value of 20.00 except DsWrocWybCon. This significant improvement in prediction accuracy highlights the effectiveness of the proposed methods in forecasting air quality for different concentrations of air pollutants. The lower MAPE values indicate that the proposed methods provide more reliable and precise predictions. Detailed MAPE values for each air pollutant from every air quality monitoring station may be seen in the Appendix as Table A.25-A.32.

4.5.3. RMSE analysis

The mean values of RMSE are shown in Table 8, both for all the proposed methods, and state-of-the-art algorithms. The mean values are calculated based on every air pollutant from each air quality station.

Similarly as in the previous cases, the proposed methods obtained lower values of RMSE than methods from the literature for all air quality stations. The values hesitate for about 0.1, however, MpZakopaSien and SlCzestoBacz obtained greater values, 0.17, and 0.31, respectively. On the other hand, state-of-the-art algorithms reached greater RMSE values, which in general exceed 0.2; the highest value (0.5) is observed for SlCzestoBacz. This also confirms that the proposed methods generate more accurate forecasts than existing

Table 6
Mean MAE values for the proposed and state-of-the-art methods from all the air quality monitoring stations.

	proposed	state-of-the-art
DsJelGorOgin	0.07	0.11
DsWrocWybCon	0.80	0.11
LuGorzKosGdy	0.10	0.16
LuSulecDudka	0.80	0.23
MpZakopaSien	0.11	0.16
PmGdaLeczkow	0.09	0.15
SlCzestoBacz	0.18	0.34
ZpKosz	0.08	0.23

Table 7
Mean MAPE values for the proposed and state-of-the-art methods from all the air quality monitoring stations.

	proposed	state-of-the-art
DsJelGorOgin	10.78	24.85
DsWrocWybCon	9.33	13.85
LuGorzKosGdy	15.71	26.21
LuSulecDudka	9.04	26.43
MpZakopaSien	30.95	74.39
PmGdaLeczkow	9.97	31.95
SlCzestoBacz	14.96	28.83
ZpKosz	11.53	72.30

algorithms. Detailed RMSE values for each air pollutant from every air quality monitoring station may be seen in the Appendix as Table A.33-A.40.

Let us graphically summarize the obtained results. The mean MAE error measure for the proposed methods is equal to 0.09, while state-of-the-art algorithms reached 0.28. The MAPE measure obtained 0.12, and 0.33 for the proposed and methods from the literature, respectively. Finally, the RMSE error measure achieved 0.1 for our methods and 0.31 for other algorithms. The summary is presented in Fig. 8.

4.6. Summary

In this subsection, we have described an extensive experimental evaluation of the proposed methods. Analysis included the statistical hypotheses verification, and also error measures analysis. The results of the hypotheses verification conducted by Kruskal-Wallis ANOVA with post-hoc analysis clearly confirmed the high accuracy of forecasts generated by the proposed methods, compared to the original data. It also showed that algorithms from the literature are less accurate.

Analysis of error measures, including the mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) also confirmed the usefulness of the proposed methods. In all cases obtained error values are smaller, compared with state-of-the-art algorithms. This also confirms the reliability of the proposed methods.

The presented analysis clearly indicates that the proposed methods may be successfully used in the case, where there is a need to estimate a concentration due to the unavailability of the original data. Such analysis may be useful to estimate the data in case it is not available, for example, due to technical issues in the air monitoring station. Predicting air pollutant concentrations is a key tool in air quality management, education, and public health protection.

5. Conclusion

In this paper, we have dealt with forecasting the concentration of air pollutants. We have proposed four deep learning-based methods for forecasting, which included temporal convolutional network (TCN), Kolmogorov-Arnold Network (KAN), fully convolutional network (FCN), and gated recurrent unit (GRU). We have generated predictions for eight air pollutants, from eight cities during a heating season in Poland. Each of the methods was used on three different configurations. Extensive experimental evaluation, combining statistical hypotheses verification and error measures for more than 740 forecast models confirmed high prediction accuracy. Moreover, experiments revealed the advantage of the proposed methods over several state-of-the-art algorithms from the literature.

In future work, we are interested in developing another method for predicting the concentration of air pollutants. In particular, methods for long-term forecasting should be discussed. Another aspect concerns the analysis of particulate matter PM_{10} and its ingredients, which is still a huge problem in Poland. Additionally, we plan to incorporate meteorological data into our models, as it plays a crucial role in accurately forecasting air quality. We also aim to explore how different meteorological factors, such as temperature, humidity, and wind speed, influence the dispersion of pollutants during summer and winter seasons. This will allow us to enhance the

Table 8
Mean RMSE values for the proposed and state-of-the-art methods from all the air quality monitoring stations.

	proposed	state-of-the-art
DsJelGorOgin	0.10	0.24
DsWrocWybCon	0.13	0.20
LuGorzKosGdy	0.13	0.19
LuSulecDudka	0.11	0.27
MpZakopaSien	0.17	0.23
PmGdaLeczkow	0.13	0.21
SlCzestoBacz	0.31	0.50
ZpKosz	0.11	0.33

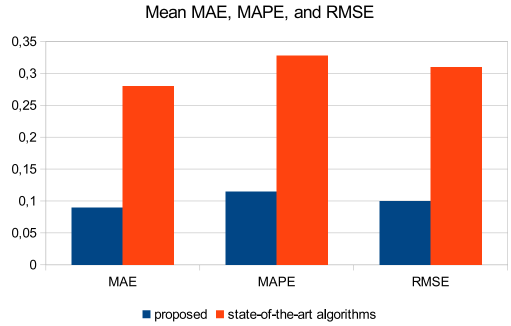


Fig. 8. Mean error values for the proposed and state-of-the-art algorithms.

robustness of our forecasting approach under varying weather conditions and improve prediction accuracy over longer periods.

CRediT authorship contribution statement

Jarosław Bernacki: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Appendix

A.1. Normality tests

Table A.9
Normality tests for all air pollutants in DsJelGorOgin air quality monitoring station.

		original	TCN 3	TCN 5	TCN 7	KAN 3	KAN 5	KAN 7	FCN 3	FCN 5	FCN 7	GRU 3	GRU 5	GRU 7	exp. sm.	exp. (Holt)	ARIMA	r. forest
C_6H_6	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	0,62	0,63	0,61	0,63	0,62	0,62	0,62	0,62	0,59	0,62	0,62	0,60	0,62	0,73	0,57	0,61	0,64
CO	h	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	1
	p	0,04	0,14	0,37	0,20	0,10	0,56	0,11	0,02	0,10	0,05	0,04	0,60	0,08	0,94	0,34	0,20	0,00
	S	0,91	0,94	0,96	0,94	0,93	0,97	0,93	0,90	0,93	0,92	0,91	0,97	0,93	0,98	0,95	0,94	0,80
NO	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,04	0,06	0,00
	S	0,64	0,61	0,66	0,62	0,63	0,63	0,63	0,64	0,66	0,64	0,64	0,66	0,63	0,88	0,91	0,92	0,59
NO_2	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1
	p	0,01	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,03	0,13	0,29	0,00
	S	0,87	0,81	0,84	0,88	0,84	0,84	0,86	0,86	0,81	0,87	0,85	0,84	0,87	0,90	0,94	0,95	0,80
NO_x	h	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,05	0,20	0,02	0,02
	S	0,78	0,83	0,85	0,83	0,78	0,78	0,78	0,78	0,80	0,78	0,78	0,78	0,77	0,92	0,94	0,90	0,89
O_3	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
	p	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,65	0,03	0,00
	S	0,80	0,85	0,84	0,90	0,79	0,70	0,79	0,80	0,77	0,80	0,82	0,80	0,81	0,90	0,97	0,91	0,71
PM_{10}	h	0	0	1	0	1	1	1	1	1	1	1	1	1	0	0	0	1
	p	0,06	0,06	0,03	0,33	0,03	0,01	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,35	0,07	0,16	0,00
	S	0,92	0,92	0,91	0,95	0,91	0,88	0,91	0,87	0,86	0,87	0,87	0,87	0,85	0,96	0,92	0,94	0,74
SO_2	h	0	0	0	0	1	1	1	1	1	1	1	0	0	1	0	0	1
	p	0,29	0,38	0,69	0,56	0,01	0,01	0,00	0,00	0,02	0,00	0,00	0,26	0,12	0,01	0,26	0,11	0,00
	S	0,95	0,96	0,97	0,97	0,87	0,89	0,86	0,82	0,90	0,81	0,83	0,95	0,93	0,87	0,95	0,93	0,81

Table A.10
Normality tests for all air pollutants in DsWrocWybCon air quality monitoring station.

		original	TCN 3	TCN 5	TCN 7	KAN 3	KAN 5	KAN 7	FCN 3	FCN 5	FCN 7	GRU 3	GRU 5	GRU 7	exp. sm.	exp. (Holt)	ARIMA	r. forest
C_6H_6	h	1	1	1	1	1	1	0	1	0	1	0	1	1	0	0	0	1
	p	0,03	0,00	0,02	0,00	0,01	0,05	0,11	0,04	0,68	0,03	0,12	0,03	0,03	0,31	0,22	0,36	0,00
	S	0,90	0,86	0,90	0,87	0,88	0,92	0,94	0,91	0,97	0,91	0,93	0,91	0,91	0,95	0,95	0,96	0,80
CO	h	1	0	1	0	1	1	1	0	1	0	0	0	0	0	1	1	1
	p	0,02	0,31	0,01	0,46	0,01	0,00	0,01	0,10	0,01	0,10	0,07	0,06	0,10	0,06	0,00	0,04	0,00
	S	0,89	0,96	0,88	0,96	0,88	0,86	0,89	0,93	0,88	0,93	0,92	0,92	0,93	0,92	0,86	0,91	0,81
NO	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,01	0,01	0,00	0,00	0,01	0,00	0,01	0,01	0,00	0,01	0,01	0,00	0,01	0,02	0,01	0,00	0,00
	S	0,89	0,89	0,86	0,82	0,87	0,85	0,87	0,89	0,83	0,89	0,89	0,84	0,89	0,90	0,89	0,86	0,73
NO_2	h	0	0	1	1	0	0	0	0	0	0	0	1	0	1	0	0	1
	p	0,26	0,06	0,01	0,01	0,25	0,05	0,15	0,19	0,06	0,19	0,24	0,04	0,24	0,01	0,48	0,37	0,01
	S	0,95	0,92	0,86	0,86	0,95	0,92	0,94	0,94	0,92	0,94	0,95	0,91	0,95	0,89	0,96	0,96	0,87
NO_x	h	0	1	1	0	0	0	0	0	1	0	0	1	1	1	0	0	1
	p	0,35	0,02	0,05	0,18	0,09	0,08	0,13	0,10	0,02	0,09	0,09	0,02	0,01	0,03	0,05	0,22	0,01
	S	0,95	0,90	0,92	0,94	0,93	0,93	0,94	0,93	0,90	0,93	0,93	0,89	0,89	0,90	0,92	0,95	0,88
O_3	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
PM_{10}	h	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	1
	p	0,01	0,09	0,11	0,06	0,01	0,00	0,01	0,01	0,00	0,02	0,00	0,00	0,02	0,07	0,01	0,28	0,01
	S	0,89	0,93	0,93	0,92	0,89	0,87	0,88	0,88	0,84	0,90	0,87	0,82	0,90	0,92	0,88	0,95	0,89
SO_2	h	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	1	0
	p	0,00	0,88	0,86	0,76	0,32	0,11	0,39	0,23	0,06	0,00	0,00	0,07	0,00	0,15	0,49	0,00	0,07
	S	0,80	0,98	0,98	0,97	0,96	0,93	0,96	0,95	0,92	0,27	0,77	0,92	0,81	0,94	0,96	0,85	0,92

Table A.11
Normality tests for all air pollutants in LuGorzKosGdy air quality monitoring station.

		original	TCN 3	TCN 5	TCN 7	KAN 3	KAN 5	KAN 7	FCN 3	FCN 5	FCN 7	GRU 3	GRU 5	GRU 7	exp. sm.	exp. (Holt)	ARIMA	r. forest
C_6H_6	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
	p	0,01	0,00	0,00	0,00	0,00	0,01	0,00	0,01	0,01	0,01	0,03	0,00	0,00	0,01	0,08	0,01	0,00
	S	0,87	0,84	0,84	0,79	0,86	0,88	0,86	0,88	0,88	0,88	0,91	0,85	0,86	0,88	0,93	0,88	0,79
CO	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	0,58	0,59	0,59	0,59	0,59	0,59	0,59	0,59	0,58	0,59	0,59	0,59	0,59	0,60	0,58	0,58	0,59
NO	h	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,09	0,19	0,40	0,00
	S	0,83	0,78	0,84	0,83	0,83	0,83	0,83	0,83	0,83	0,83	0,83	0,83	0,83	0,93	0,94	0,96	0,82
NO_2	h	0	0	0	0	1	1	1	1	1	1	1	1	1	0	1	0	0
	p	0,47	0,51	0,44	0,74	0,01	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,05	0,02	0,30	0,19
	S	0,96	0,96	0,96	0,97	0,87	0,85	0,87	0,75	0,67	0,72	0,19	0,85	0,18	0,92	0,90	0,95	0,94
NO_x	h	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1
	p	0,09	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,60	0,04	0,27	0,00
	S	0,93	0,82	0,86	0,88	0,78	0,77	0,81	0,18	0,00	0,18	0,18	0,04	0,18	0,97	0,91	0,95	0,81
O_3	h	0	0	0	0	1	1	1	0	1	0	1	1	1	0	0	0	0
	p	0,37	0,97	0,45	0,34	0,03	0,01	0,01	0,10	0,00	0,07	0,05	0,01	0,05	0,10	0,06	0,36	0,65
	S	0,96	0,99	0,96	0,96	0,91	0,86	0,87	0,93	0,85	0,92	0,92	0,86	0,92	0,93	0,92	0,96	0,97
PM_{10}	h	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,35	0,00	0,00	0,00
	S	0,68	0,71	0,78	0,75	0,68	0,65	0,69	0,70	0,65	0,69	0,69	0,70	0,69	0,96	0,75	0,68	0,63
SO_2	h	0	1	0	0	1	1	1	1	1	1	1	1	1	0	0	0	1
	p	0,97	0,03	0,12	0,37	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,16	0,38	0,07	0,00
	S	0,99	0,91	0,93	0,96	0,77	0,90	0,78	0,82	0,87	0,76	0,76	0,89	0,75	0,94	0,96	0,92	0,84

Table A.12
Normality tests for all air pollutants in LuSulecDudka air quality monitoring station.

		original	TCN 3	TCN 5	TCN 7	KAN 3	KAN 5	KAN 7	FCN 3	FCN 5	FCN 7	GRU 3	GRU 5	GRU 7	exp. sm.	exp. (Holt)	ARIMA	r. forest
CO	h	0	0	0	0	1	1	1	0	0	1	1	0	1	1	0	0	1
	p	0,39	0,45	0,12	0,22	0,01	0,00	0,01	0,07	0,32	0,03	0,01	0,31	0,01	0,00	0,44	0,76	0,00
	S	0,96	0,96	0,93	0,95	0,87	0,85	0,88	0,92	0,95	0,91	0,89	0,95	0,88	0,70	0,96	0,97	0,81
NO	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	0,85	0,83	0,83	0,87	0,84	0,84	0,84	0,84	0,84	0,84	0,85	0,84	0,85	0,57	0,51	0,74	0,78
NO_2	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
NO_x	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

(*continued on next page*)

Table A.12 (*continued*)

		original	TCN 3	TCN 5	TCN 7	KAN 3	KAN 5	KAN 7	FCN 3	FCN 5	FCN 7	GRU 3	GRU 5	GRU 7	exp. sm.	exp. (Holt)	ARIMA	r. forest
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
O_3	h	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1
	p	0,08	0,73	0,84	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,12	0,00
	S	0,93	0,97	0,98	0,90	0,18	0,79	0,19	0,18	0,02	0,18	0,18	0,02	0,18	0,84	0,18	0,93	0,78
PM_{10}	h	0	1	0	0	1	0	1	1	1	1	1	1	1	0	0	0	0
	p	0,21	0,02	0,22	0,67	0,00	0,05	0,01	0,00	0,04	0,00	0,00	0,00	0,00	0,50	0,26	0,38	0,18
	S	0,94	0,89	0,95	0,97	0,84	0,92	0,88	0,22	0,91	0,19	0,18	0,51	0,18	0,96	0,95	0,96	0,94
SO_2	h	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1
	p	0,46	0,33	0,04	0,09	0,10	0,48	0,17	0,14	0,59	0,31	0,24	0,53	0,18	0,02	0,00	0,19	0,00
	S	0,96	0,96	0,91	0,93	0,93	0,96	0,94	0,94	0,97	0,96	0,95	0,96	0,94	0,90	0,18	0,94	0,84

Table A.13
Normality tests for all air pollutants in MpZakopaSien air quality monitoring station.

		original	TCN 3	TCN 5	TCN 7	KAN 3	KAN 5	KAN 7	FCN 3	FCN 5	FCN 7	GRU 3	GRU 5	GRU 7	exp. sm.	exp. (Holt)	ARIMA	r. forest
CO	h	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1
	p	0,01	0,08	0,02	0,03	0,00	0,02	0,00	0,01	0,01	0,01	0,01	0,02	0,01	0,12	0,02	0,12	0,00
	S	0,88	0,93	0,90	0,91	0,87	0,90	0,87	0,88	0,89	0,88	0,88	0,89	0,88	0,93	0,90	0,94	0,80
NO	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	0,80	0,96	0,98	0,90	0,84	0,84	0,83	0,81	0,82	0,84	0,80	0,84	0,81	0,94	0,92	0,67	1,00
NO_2	h	0	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	1
	p	0,08	0,02	0,04	0,28	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,19	0,14	0,39	0,00
	S	0,93	0,90	0,91	0,95	0,82	0,80	0,80	0,80	0,74	0,79	0,19	0,04	0,18	0,94	0,94	0,96	0,76
NO_x	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,96	0,92	1,00
O_3	h	1	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	1
	p	0,01	0,14	0,34	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,09	0,13	0,15	0,00
	S	0,87	0,94	0,95	0,87	0,83	0,84	0,84	0,80	0,73	0,81	0,73	0,70	0,80	0,93	0,93	0,94	0,76
PM_{10}	h	1	1	1	1	0	1	0	1	1	1	1	1	1	0	1	1	1
	p	0,00	0,00	0,00	0,01	0,05	0,03	0,09	0,03	0,00	0,02	0,01	0,00	0,01	0,12	0,00	0,03	0,00
	S	0,81	0,84	0,78	0,87	0,92	0,90	0,93	0,91	0,87	0,90	0,89	0,84	0,87	0,94	0,82	0,91	0,80
SO_2	h	1	1	0	0	1	0	1	1	0	1	1	0	1	0	0	0	1
	p	0,04	0,00	0,32	0,10	0,05	0,08	0,04	0,04	0,09	0,04	0,04	0,08	0,04	0,24	0,41	0,07	0,04
	S	0,91	0,76	0,96	0,93	0,92	0,93	0,91	0,91	0,93	0,91	0,91	0,93	0,91	0,95	0,96	0,92	0,91

Table A.14
Normality tests for all air pollutants in PmGdaLeczkow air quality monitoring station.

		original	TCN 3	TCN 5	TCN 7	KAN 3	KAN 5	KAN 7	FCN 3	FCN 5	FCN 7	GRU 3	GRU 5	GRU 7	exp. sm.	exp. (Holt)	ARIMA	r. forest
C_6H_6	h	1	0	1	1	0	1	0	1	1	1	1	1	1	0	0	1	0
	p	0,02	0,12	0,00	0,04	0,07	0,02	0,06	0,02	0,02	0,02	0,02	0,03	0,02	0,14	0,81	0,00	0,13
	S	0,90	0,94	0,85	0,91	0,92	0,89	0,92	0,90	0,89	0,90	0,90	0,91	0,90	0,94	0,98	0,86	0,94
CO	h	0	0	0	0	1	1	1	0	0	0	0	0	1	1	0	1	0
	p	0,05	0,30	0,62	0,23	0,00	0,00	0,00	0,08	0,24	0,13	0,12	0,07	0,00	0,01	0,37	0,02	0,06
	S	0,92	0,96	0,97	0,95	0,62	0,70	0,70	0,93	0,95	0,94	0,94	0,93	0,62	0,88	0,96	0,90	0,92
NO	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,88	0,08	0,00
	S	0,86	0,86	0,86	0,83	0,86	0,86	0,86	0,86	0,84	0,85	0,85	0,86	0,85	0,90	0,98	0,93	0,85
NO_2	h	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
	p	0,00	0,09	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,00	0,00	0,00	0,03	0,26	0,00	0,00
	S	0,81	0,93	0,89	0,89	0,87	0,87	0,87	0,88	0,88	0,87	0,86	0,85	0,87	0,91	0,95	0,75	0,62
NO_x	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,96	1,00	1,00
O_3	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
PM_{10}	h	0	0	0	0	1	1	0	1	1	1	1	1	1	0	0	0	1
	p	0,57	0,84	0,26	0,92	0,01	0,00	0,05	0,03	0,00	0,03	0,00	0,00	0,00	0,23	0,49	0,89	0,00
	S	0,97	0,98	0,95	0,98	0,89	0,87	0,92	0,91	0,67	0,90	0,79	0,64	0,85	0,95	0,96	0,98	0,76
SO_2	h	0	0	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1
	p	0,17	0,24	0,18	0,00	0,03	0,00	0,02	0,02	0,01	0,03	0,02	0,26	0,06	0,00	0,00	0,00	0,00
	S	0,94	0,95	0,94	0,85	0,90	0,85	0,90	0,89	0,87	0,91	0,90	0,95	0,92	0,84	0,85	0,74	0,82

Table A.15
Normality tests for all air pollutants in SlCzestoBacz air quality monitoring station.

		original	TCN 3	TCN 5	TCN 7	KAN 3	KAN 5	KAN 7	FCN 3	FCN 5	FCN 7	GRU 3	GRU 5	GRU 7	exp. sm.	exp. (Holt)	ARIMA	r. forest
C_6H_6	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	0,68	0,70	0,66	0,69	0,68	0,68	0,68	0,68	0,68	0,68	0,68	0,68	0,68	0,68	0,75	0,69	0,71
CO	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	0,59	0,59	0,57	0,59	0,59	0,61	0,59	0,59	0,59	0,59	0,59	0,59	0,59	0,59	0,59	0,59	0,59
NO	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	0,77	0,79	0,80	0,77	0,77	0,81	0,78	0,77	0,78	0,77	0,77	0,79	0,77	0,80	0,75	0,78	0,83
NO_2	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	1,00	0,96	1,00	0,98	1,00	0,98	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,96	0,96	0,87	0,89
NO_x	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
O_3	h	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,20	0,77	0,62	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	0,94	0,97	0,97	0,99	0,19	0,85	0,78	0,18	0,04	0,18	0,18	0,00	0,18	0,80	0,82	0,77	0,68
PM_{10}	h	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1
	p	0,91	0,12	0,09	0,10	0,01	0,03	0,04	0,02	0,01	0,04	0,02	0,01	0,00	0,00	0,04	0,70	0,00
	S	0,98	0,93	0,93	0,93	0,88	0,91	0,91	0,90	0,87	0,91	0,90	0,88	0,85	0,80	0,91	0,97	0,85
SO_2	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
	p	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,03	0,14
	S	0,82	0,74	0,79	0,87	0,77	0,76	0,80	0,80	0,78	0,82	0,82	0,82	0,80	0,68	0,88	0,91	0,94

Table A.16
Normality tests for all air pollutants in ZpKosz air quality monitoring station.

		original	TCN 3	TCN 5	TCN 7	KAN 3	KAN 5	KAN 7	FCN 3	FCN 5	FCN 7	GRU 3	GRU 5	GRU 7	exp. sm.	exp. (Holt)	ARIMA	r. forest
C_6H_6	h	0	1	0	0	0	1	0	0	0	0	0	0	0	1	1	1	1
	p	0,10	0,01	0,09	0,07	0,18	0,02	0,16	0,10	0,31	0,12	0,13	0,09	0,24	0,01	0,00	0,01	0,00
	S	0,93	0,86	0,93	0,92	0,94	0,90	0,94	0,93	0,96	0,94	0,94	0,93	0,95	0,87	0,84	0,89	0,83
CO	h	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1
	p	0,57	0,01	0,00	0,00	0,01	0,02	0,00	0,01	0,02	0,01	0,01	0,18	0,00	0,03	0,23	0,71	0,00
	S	0,97	0,88	0,87	0,81	0,88	0,89	0,73	0,88	0,90	0,88	0,88	0,94	0,86	0,91	0,95	0,97	0,83
NO	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	0,69	0,75	0,72	0,75	0,72	0,70	0,71	0,73	0,70	0,74	0,75	0,70	0,75	0,70	0,68	0,64	0,77
NO_2	h	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
	p	0,04	0,13	0,01	0,02	0,04	0,04	0,04	0,04	0,03	0,04	0,04	0,04	0,04	0,04	0,37	0,01	0,03
	S	0,91	0,94	0,87	0,89	0,91	0,91	0,91	0,91	0,91	0,91	0,91	0,91	0,91	0,91	0,96	0,87	0,91
NO_x	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,94	1,00	1,00
O_3	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
PM_{10}	h	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	p	0,43	0,99	0,05	0,59	0,52	0,29	0,41	0,17	0,14	0,23	0,33	0,10	0,30	0,07	0,14	0,01	0,18
	S	0,96	0,99	0,92	0,97	0,96	0,95	0,96	0,94	0,94	0,95	0,95	0,93	0,95	0,93	0,94	0,87	0,94
SO_2	h	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	p	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	S	0,71	0,81	0,75	0,71	0,71	0,77	0,71	0,71	0,70	0,71	0,71	0,71	0,71	0,83	0,75	0,69	0,83

A.2. Error measures – preliminaries

Symbol markings:

- "min" stands for the minimal value of the concentration of the selected air pollutant from the data from GIOŚ;
- "max" stands for the maximum value of the concentration of the selected air pollutant from the data from GIOŚ;
- A number next to the name of the proposed method denotes the size of the window (e.g. TCN 3 denotes the TCN with a window of size 3; FCN 7 denotes the FCN with a window of size 7);
- mean$^{(1)}$ stands for a mean obtained for all air pollutants for a selected method (horizontal row);
- mean$^{(2)}$ denotes the general mean both for the proposed methods and for state-of-the-art methods (in fact it is a mean of mean$^{(1)}$, both for the proposed and state-of-the-art methods);

- *exp*.sm. denotes exponential smoothing algorithm; *exp*. (Holt) denotes Holt-Winters exponential smoothing algorithm; ARIMA stands for autoregressive integrated moving average algorithm; r.forest denotes the random forest algorithm.

This information concerns all the tables presented in Subsec. Appendix A.2.1. Appendix A.2.2. and Appendix A.2.3.

A.2.1. Full MAE results

Table A.17
MAE values for DsJelGorOgin air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean$^{(1)}$	mean$^{(2)}$
min	0.30	0.25	0.00	5.50	6.33	10.87	9.43	3.17	–	–
max	2.86	0.63	9.59	24.85	30.51	54.82	46.26	8.73	–	–
TCN 3	0.21	0.07	0.29	1.87	2.67	3.38	4.17	0.78	1.68	0.07
TCN 5	0.21	0.07	0.63	1.92	3.09	4.44	5.97	0.83	2.14	
TCN 7	0.20	0.06	0.46	1.82	3.13	5.59	4.89	0.86	2.13	
KAN 3	0.00	0.03	0.05	0.31	0.24	1.83	1.55	0.63	0.58	
KAN 5	0.07	0.06	0.13	0.97	2.34	3.96	4.23	0.70	1.56	
KAN 7	0.00	0.03	0.04	0.17	0.26	2.15	1.55	0.65	0.61	
FCN 3	0.00	0.02	0.00	0.20	0.16	2.38	2.88	0.58	0.78	
FCN 5	0.02	0.06	0.08	0.98	1.08	3.93	3.43	0.60	1.27	
FCN 7	0.00	0.02	0.02	0.06	0.20	2.32	2.83	0.58	0.75	
GRU 3	0.00	0.03	0.02	0.41	0.23	2.80	2.74	0.50	0.84	
GRU 5	0.03	0.06	0.09	0.62	1.21	3.24	3.02	0.27	1.07	
GRU 7	0.00	0.03	0.04	0.06	0.41	2.91	3.11	0.17	0.84	
exp. sm.	0.45	0.08	1.54	3.22	5.26	7.84	7.41	0.77	3.32	0.11
exp. (Holt)	0.24	0.04	0.96	2.33	2.73	3.80	3.42	0.57	1.76	
ARIMA	0.14	0.03	0.48	1.45	1.92	2.65	2.16	0.39	1.15	
r. forest	0.19	0.04	0.34	1.34	2.00	2.46	4.37	0.48	1.40	

Table A.18
MAE values for DsWrocWybCon air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean$^{(1)}$	mean$^{(2)}$
min	0.28	0.22	0.33	6.47	7.64	21.40	9.97	4.56	–	–
max	0.98	0.48	3.58	27.80	31.39	58.36	25.32	7.85	–	–
TCN 3	0.07	0.04	0.32	2.76	3.93	3.93	3.05	0.62	1.84	0.08
TCN 5	0.07	0.03	0.31	2.80	4.30	4.30	3.32	0.60	1.97	
TCN 7	0.09	0.04	0.33	3.09	3.83	3.83	3.19	0.68	1.88	
KAN 3	0.09	0.02	0.03	0.68	0.80	0.80	0.04	0.31	0.35	
KAN 5	0.11	0.03	0.16	1.83	1.93	1.93	1.74	0.30	1.00	
KAN 7	0.04	0.02	0.02	0.92	0.53	0.53	1.40	0.31	0.47	
FCN 3	0.02	0.02	0.00	0.48	0.47	0.47	0.58	0.30	0.29	
FCN 5	0.08	0.03	0.06	1.54	1.60	1.60	1.66	0.31	0.86	
FCN 7	0.01	0.02	0.00	0.48	0.52	0.52	0.46	0.40	0.30	
GRU 3	0.03	0.02	0.01	0.31	0.53	0.53	1.00	0.36	0.35	
GRU 5	0.01	0.02	0.05	0.94	1.56	1.56	1.44	0.30	0.74	
GRU 7	0.01	0.02	0.00	0.35	1.04	1.04	1.74	0.27	0.56	
exp. sm.	0.13	0.05	0.46	3.88	4.55	4.95	3.12	0.34	2.18	0.11
exp. (Holt)	0.09	0.02	0.28	2.92	3.17	3.95	1.84	0.31	1.57	
ARIMA	0.04	0.02	0.20	1.76	1.88	3.13	1.53	0.10	1.08	
r. forest	0.06	0.01	0.20	2.60	2.11	3.16	1.16	0.16	1.18	

Table A.19
MAE values for LuGorzKosGdy air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean$^{(1)}$	mean$^{(2)}$
min	0.01	0.20	0.51	14.50	20.65	1.89	7.84	4.20	–	–
max	0.26	0.60	17.33	48.85	69.75	45.00	37.09	8.44	–	–
TCN 3	0.02	0.03	1.27	4.96	8.18	4.68	1.85	0.94	2.74	0.10
TCN 5	0.02	0.04	2.31	8.03	9.42	4.98	2.17	0.85	3.48	
TCN 7	0.03	0.03	1.15	5.22	7.03	6.22	2.27	0.70	2.83	
KAN 3	0.02	0.02	0.00	5.09	6.58	1.95	0.63	0.49	1.85	
KAN 5	0.03	0.03	0.18	4.41	7.05	2.86	1.07	0.60	2.03	
KAN 7	0.03	0.03	0.06	5.04	6.32	2.56	0.44	0.49	1.87	
FCN 3	0.03	0.02	0.00	3.75	8.31	1.72	0.48	0.54	1.85	
FCN 5	0.03	0.03	0.05	5.50	9.06	2.68	1.22	0.57	2.39	
FCN 7	0.03	0.02	0.00	3.67	8.31	1.61	0.53	0.48	1.83	
GRU 3	0.01	0.01	0.00	7.62	8.31	1.47	0.39	0.70	2.31	

(continued on next page)

Table A.19 (*continued*)

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
GRU 5	0.02	0.03	0.00	7.52	9.07	2.50	0.41	0.57	2.51	
GRU 7	0.03	0.02	0.00	7.63	8.30	1.51	0.40	0.47	2.29	
exp. sm.	0.03	0.06	5.58	9.50	10.40	10.15	12.19	1.32	6.15	0.16
exp. (Holt)	0.02	0.03	5.60	10.21	9.37	2.88	1.96	0.50	3.82	
ARIMA	0.01	0.01	1.89	10.54	4.73	0.96	0.97	0.41	2.44	
r. forest	0.02	0.03	1.47	3.64	5.65	14.08	1.38	0.37	3.33	

Table A.20
MAE values for LuSulecDudka air quality monitoring station.

	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.42	3.50	5.95	11.20	22.75	8.67	2.77	–	–
max	1.09	19.80	33.66	63.36	30.66	63.24	4.90	–	–
TCN 3	0.04	1.35	2.04	3.33	3.04	2.62	0.28	1.81	0.08
TCN 5	0.05	1.39	2.74	3.95	3.82	2.58	0.36	2.13	
TCN 7	0.04	0.99	2.28	2.93	3.70	2.72	0.36	1.86	
KAN 3	0.03	0.13	0.25	2.26	1.31	2.43	0.23	0.95	
KAN 5	0.04	0.26	0.42	2.40	1.53	2.66	0.26	1.08	
KAN 7	0.03	0.14	0.27	2.42	1.31	2.58	0.18	0.99	
FCN 3	0.03	0.18	0.48	3.06	1.30	2.82	0.22	1.16	
FCN 5	0.03	0.22	0.45	3.42	1.53	2.76	0.26	1.24	
FCN 7	0.03	0.20	0.41	2.79	1.30	2.83	0.17	1.10	
GRU 3	0.02	0.10	0.29	2.50	1.31	2.83	0.17	1.03	
GRU 5	0.03	0.17	0.28	2.77	1.53	2.89	0.26	1.13	
GRU 7	0.03	0.06	0.29	2.92	1.31	2.83	0.23	1.10	
exp. sm.	0.31	9.65	4.03	33.59	1.02	23.14	0.31	10.29	0.23
exp. (Holt)	0.03	8.57	15.80	1.16	4.83	3.97	1.09	5.06	
ARIMA	0.02	6.68	1.04	1.93	0.21	1.84	0.77	1.79	
r. forest	0.02	6.06	1.16	2.10	0.64	2.18	0.24	1.77	

Table A.21
MAE values for MpZakopaSien air quality monitoring station.

	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.24	0.85	6.32	7.70	1.83	5.26	0.51	–	–
max	1.66	124.44	58.62	248.70	62.12	102.34	14.86	–	–
TCN 3	0.14	15.52	8.68	28.61	6.30	11.64	1.99	10.41	0.11
TCN 5	0.17	16.46	10.32	35.16	6.88	13.64	1.67	12.04	
TCN 7	0.19	12.37	11.17	25.08	7.86	12.11	1.87	10.09	
KAN 3	0.02	13.41	7.10	44.03	1.53	9.82	0.37	10.90	
KAN 5	0.11	13.61	7.98	48.86	4.90	11.03	0.67	12.45	
KAN 7	0.02	13.30	7.28	42.99	1.79	9.98	0.16	10.79	
FCN 3	0.00	12.97	10.51	57.19	1.82	8.16	0.01	12.95	
FCN 5	0.05	12.03	9.12	58.03	4.35	9.89	0.29	13.39	
FCN 7	0.00	13.24	8.14	43.81	1.86	8.19	0.01	10.75	
GRU 3	0.03	12.94	13.63	57.28	3.82	8.36	0.01	13.72	
GRU 5	0.09	15.74	13.88	58.26	4.15	9.67	0.44	14.60	
GRU 7	0.00	12.81	13.63	57.25	3.86	9.14	0.03	13.82	
exp. sm.	0.45	20.94	14.41	48.03	12.29	16.26	2.03	16.35	0.16
exp. (Holt)	0.14	35.62	6.91	29.77	7.04	8.71	1.77	12.85	
ARIMA	0.46	8.26	12.45	15.77	6.96	7.81	0.95	7.52	
r. forest	0.22	14.12	4.58	44.18	9.78	5.80	6.54	12.17	

Table A.22
MAE values for PmGdaLeczkow air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.23	0.42	0.48	3.97	4.79	20.93	4.02	2.16	–	–
max	1.24	0.56	2.93	25.81	30.30	66.57	15.45	5.22	–	–
TCN 3	0.08	0.03	0.23	1.81	2.50	6.42	2.28	0.34	1.71	0.09
TCN 5	0.08	0.03	0.16	2.84	3.25	9.79	2.22	0.54	2.36	
TCN 7	0.10	0.04	0.17	2.45	2.66	6.91	2.12	0.33	1.85	
KAN 3	0.06	0.03	0.04	0.37	0.64	7.43	1.74	0.25	1.32	
KAN 5	0.06	0.03	0.08	1.40	1.72	7.75	1.81	0.26	1.64	
KAN 7	0.06	0.03	0.01	0.60	0.74	7.42	1.65	0.24	1.34	
FCN 3	0.00	0.02	0.00	0.35	0.50	7.51	0.27	0.17	1.10	
FCN 5	0.06	0.02	0.05	0.43	0.60	7.88	1.45	0.26	1.34	
FCN 7	0.00	0.02	0.00	0.24	0.50	7.57	0.34	0.23	1.11	
GRU 3	0.00	0.02	0.01	0.18	0.24	7.48	1.32	0.22	1.18	

(*continued on next page*)

Table A.22 (*continued*)

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
GRU 5	0.04	0.02	0.02	0.14	0.67	7.80	1.76	0.19	1.33	
GRU 7	0.00	0.03	0.01	0.57	0.24	7.46	0.67	0.12	1.14	
exp. sm.	0.15	0.02	0.30	3.51	4.12	9.99	1.80	0.74	2.58	0.15
exp. (Holt)	0.13	0.01	0.30	3.90	3.73	10.04	1.45	0.69	2.53	
ARIMA	0.09	0.01	0.19	2.59	3.59	11.55	0.89	0.56	2.43	
r. forest	0.13	0.01	0.20	2.85	3.72	9.76	1.32	0.68	2.33	

Table A.23
MAE values for SlCzestoBacz air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.37	0.21	0.72	4.39	5.49	25.00	7.74	2.57	–	–
max	2.38	0.42	1.99	17.56	19.44	53.58	26.66	19.64	–	–
TCN 3	0.12	0.02	0.20	13.79	1.64	5.02	3.35	1.69	3.23	0.18
TCN 5	0.11	0.03	0.18	14.98	1.48	6.21	3.56	2.38	3.62	
TCN 7	0.09	0.03	0.23	16.27	1.51	5.55	4.44	1.99	3.77	
KAN 3	0.00	0.02	0.07	14.46	0.32	5.48	2.19	0.52	2.88	
KAN 5	0.06	0.03	0.18	17.10	0.89	5.39	2.15	1.04	3.35	
KAN 7	0.01	0.02	0.02	12.25	0.42	5.19	1.88	0.40	2.52	
FCN 3	0.00	0.02	0.01	4.65	0.21	5.62	1.14	0.30	1.49	
FCN 5	0.04	0.02	0.10	15.71	0.71	5.67	1.75	0.61	3.08	
FCN 7	0.01	0.02	0.03	10.46	0.17	5.59	0.96	0.10	2.17	
GRU 3	0.01	0.01	0.00	3.63	0.22	5.57	1.06	0.04	1.32	
GRU 5	0.01	0.02	0.06	12.77	0.18	5.64	1.71	0.09	2.56	
GRU 7	0.00	0.02	0.00	2.08	0.13	5.59	3.56	0.39	1.47	
exp. sm.	0.42	0.03	0.19	11.56	3.47	11.55	5.53	4.01	4.59	0.34
exp. (Holt)	0.64	0.02	0.19	16.90	4.00	8.95	5.90	1.71	4.79	
ARIMA	0.31	0.02	0.05	17.52	3.32	8.68	4.50	0.80	4.40	
r. forest	0.50	0.01	0.13	18.58	3.95	8.92	5.34	5.75	5.40	

Table A.24
MAE values for ZpKosz air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.38	0.06	0.45	3.03	4.19	30.63	5.19	0.18	–	–
max	1.50	0.16	13.39	19.92	38.06	74.27	14.58	7.77	–	–
TCN 3	0.11	0.01	0.71	1.61	3.83	5.97	2.38	0.75	1.92	0.08
TCN 5	0.12	0.02	0.82	2.13	4.03	7.29	2.11	0.77	2.16	
TCN 7	0.10	0.02	0.91	2.03	4.21	4.78	1.69	0.81	1.82	
KAN 3	0.06	0.02	0.12	0.12	1.15	6.82	1.15	0.03	1.18	
KAN 5	0.08	0.02	0.17	0.66	1.82	6.96	0.89	0.45	1.38	
KAN 7	0.06	0.02	0.08	0.08	0.75	6.78	0.38	0.01	1.02	
FCN 3	0.04	0.01	0.07	0.11	0.77	6.95	0.67	0.01	1.08	
FCN 5	0.07	0.02	0.06	0.36	1.81	7.11	0.95	0.06	1.31	
FCN 7	0.01	0.02	0.07	0.09	0.79	6.92	0.14	0.01	1.01	
GRU 3	0.01	0.02	0.08	0.05	1.18	6.92	0.64	0.11	1.12	
GRU 5	0.07	0.02	0.08	0.11	1.69	7.06	0.36	0.02	1.18	
GRU 7	0.05	0.01	0.11	0.04	1.16	6.92	0.12	0.02	1.05	
exp. sm.	0.32	0.02	2.08	2.94	8.27	9.91	4.20	2.64	3.80	0.23
exp. (Holt)	0.35	0.02	1.95	3.20	8.46	9.75	2.78	1.57	3.51	
ARIMA	0.24	0.01	1.80	2.11	7.12	8.65	3.86	1.76	3.19	
r. forest	0.36	0.02	1.87	2.95	7.48	7.64	2.99	2.08	3.17	

A.2.2. Full MAPE results

Table A.25
MAPE values for DsJelGorOgin air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.30	0.25	0.00	5.50	6.33	10.87	9.43	3.17	–	–
max	2.86	0.63	9.59	24.85	30.51	54.82	46.26	8.73	–	–
TCN 3	19.75	19.79	30.24	21.75	22.09	15.71	17.32	13.35	20.00	10.78
TCN 5	23.64	14.94	39.78	18.72	24.59	18.58	24.60	15.48	22.54	
TCN 7	21.91	14.46	31.13	17.66	28.11	21.62	21.82	14.41	21.39	
KAN 3	0.37	7.69	12.02	3.55	1.45	5.74	5.05	9.89	5.72	
KAN 5	7.71	12.42	22.96	9.42	15.55	17.36	15.53	11.27	14.03	

(*continued on next page*)

Table A.25 (*continued*)

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
KAN 7	0.00	7.67	9.91	2.01	1.91	7.58	5.49	10.20	5.60	
FCN 3	0.17	5.49	6.23	1.66	0.64	6.38	8.94	9.20	4.64	
FCN 5	3.42	14.44	17.50	8.41	6.38	14.85	12.04	9.66	10.84	
FCN 7	0.55	5.48	6.49	0.43	0.74	6.26	8.73	9.16	4.73	
GRU 3	0.01	6.96	7.43	2.42	1.12	7.86	8.29	7.66	5.22	
GRU 5	4.73	12.60	19.81	4.31	5.62	10.82	9.80	4.88	9.07	
GRU 7	0.00	8.00	8.66	0.34	1.49	8.38	9.66	2.86	4.93	
exp. sm.	42.97	19.08	70.29	28.05	35.74	29.74	30.79	13.93	33.82	24.85
exp. (Holt)	37.00	8.67	113.94	26.07	27.43	20.85	16.08	9.96	32.50	
ARIMA	22.56	8.19	57.18	15.84	17.92	24.61	16.70	6.59	16.06	
r. forest	16.83	9.48	34.53	13.52	15.60	11.09	19.42	7.90	16.05	

Table A.26
MAPE values for DsWrocWybCon air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.28	0.22	0.33	6.47	7.64	21.40	9.97	4.56	–	–
max	0.98	0.48	3.58	27.80	31.39	58.36	25.32	7.85	–	–
TCN 3	11.94	13.11	23.30	21.42	26.56	26.56	21.11	10.73	19.34	9.33
TCN 5	13.33	9.73	24.40	21.15	32.10	32.10	21.66	11.11	20.70	
TCN 7	14.95	11.79	22.99	22.17	27.00	27.00	23.88	12.05	20.23	
KAN 3	16.11	7.41	3.57	3.68	4.15	4.15	0.27	5.32	5.58	
KAN 5	20.15	8.84	13.21	11.84	10.90	10.90	10.28	5.21	11.42	
KAN 7	7.49	7.53	2.85	5.35	2.52	2.52	8.34	5.31	5.24	
FCN 3	4.54	7.05	0.15	2.41	2.04	2.04	3.26	5.27	3.35	
FCN 5	14.18	7.89	6.56	9.66	8.69	8.69	9.74	5.33	8.84	
FCN 7	2.15	7.05	0.26	2.58	2.25	2.25	2.33	7.01	3.24	
GRU 3	5.29	6.23	1.00	1.36	2.28	2.28	5.21	6.27	3.74	
GRU 5	2.14	7.41	5.63	4.50	7.91	7.91	7.96	5.11	6.07	
GRU 7	0.92	6.84	0.26	1.61	4.34	4.34	10.33	4.68	4.17	
exp. sm.	24.77	14.20	32.22	26.50	27.30	13.00	18.77	6.13	20.36	13.85
exp. (Holt)	18.13	6.92	25.04	23.57	21.89	11.44	12.30	5.34	15.58	
ARIMA	10.71	5.72	12.21	14.49	13.75	9.12	9.41	1.66	9.64	
r. forest	11.14	4.49	13.54	17.73	13.22	8.89	6.96	2.67	9.83	

Table A.27
MAPE values for LuGorzKosGdy air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.01	0.20	0.51	14.50	20.65	1.89	7.84	4.20	–	–
max	0.26	0.60	17.33	48.85	69.75	45.00	37.09	8.44	–	–
TCN 3	31.15	9.97	35.08	19.20	25.64	21.94	17.71	15.42	22.02	15.71
TCN 5	31.54	13.80	45.94	33.31	24.51	25.79	15.96	13.79	25.58	
TCN 7	40.72	10.02	24.22	19.76	20.81	25.07	16.13	11.29	21.00	
KAN 3	36.99	7.04	0.17	18.12	19.07	7.48	4.22	7.53	12.58	
KAN 5	36.55	10.29	6.49	15.00	19.88	11.31	7.25	9.17	14.49	
KAN 7	37.23	8.83	1.12	17.94	18.22	9.65	2.26	7.55	12.85	
FCN 3	43.23	8.02	0.01	12.62	23.44	6.92	2.50	8.18	13.11	
FCN 5	36.08	10.56	3.00	20.09	25.04	11.04	7.60	8.84	15.28	
FCN 7	36.41	7.72	0.00	11.77	23.45	6.36	3.04	7.37	12.01	
GRU 3	19.84	3.63	0.00	26.19	23.44	5.47	1.66	10.27	11.31	
GRU 5	34.27	10.28	0.05	26.57	24.98	9.43	2.14	8.78	14.56	
GRU 7	37.42	7.30	0.00	26.21	23.51	5.90	1.72	7.23	13.66	
exp. sm.	44.29	16.07	59.93	41.21	25.11	71.22	49.86	24.93	41.58	26.21
exp. (Holt)	33.61	11.10	60.17	28.05	29.28	34.14	17.53	7.79	27.71	
ARIMA	15.36	3.40	53.12	28.17	12.02	8.35	7.55	6.47	16.80	
r. forest	25.70	8.94	32.18	12.42	15.14	40.69	9.23	5.79	18.76	

Table A.28
MAPE values for LuSulecDudka air quality monitoring station.

	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.42	3.50	5.95	11.20	22.75	8.67	2.77	–	–
max	1.09	19.80	33.66	63.36	30.66	63.24	4.90	–	–
TCN 3	8.32	19.18	15.57	15.29	11.11	18.23	7.80	13.64	9.04
TCN 5	10.38	16.04	15.89	19.94	14.83	19.28	10.16	15.22	
TCN 7	8.52	15.19	19.79	12.76	13.47	20.89	11.24	14.55	
KAN 3	6.46	1.79	1.69	6.97	5.12	17.48	6.75	6.61	
KAN 5	7.16	4.41	3.46	7.50	5.92	19.20	7.49	7.88	

(*continued on next page*)

Table A.28 (*continued*)

	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
KAN 7	5.19	1.95	1.78	7.94	5.12	18.73	5.29	6.57	
FCN 3	5.38	2.69	3.61	9.27	5.11	20.65	6.39	7.58	
FCN 5	6.11	3.86	3.61	10.81	5.92	20.22	7.37	8.27	
FCN 7	5.12	3.24	3.38	8.36	5.11	20.65	4.88	7.25	
GRU 3	4.97	1.54	1.58	7.05	5.11	20.65	4.63	6.51	
GRU 5	5.78	3.05	1.68	7.32	5.92	20.82	7.32	7.41	
GRU 7	5.34	0.75	1.65	8.76	5.11	20.65	6.67	6.99	
exp. sm.	31.60	68.37	22.61	65.57	3.77	59.02	8.35	37.04	26.43
exp. (Holt)	5.82	73.04	80.81	5.63	15.91	41.72	23.42	35.20	
ARIMA	3.54	70.38	9.00	8.85	0.85	19.93	17.96	18.64	
r. forest	3.42	57.43	8.99	8.57	2.54	16.04	6.95	14.85	

Table A.29
MAPE values for MpZakopaSien air quality monitoring station.

	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.24	0.85	6.32	7.70	1.83	5.26	0.51	–	–
max	1.66	124.44	58.62	248.70	62.12	102.34	14.86	–	–
TCN 3	19.47	67.03	32.03	51.82	38.37	38.11	27.71	39.22	30.95
TCN 5	19.98	73.99	32.14	51.50	39.05	48.35	23.66	41.24	
TCN 7	22.58	60.61	40.68	36.38	41.78	38.88	26.04	38.14	
KAN 3	2.72	31.73	19.44	74.58	7.06	26.92	4.46	23.84	
KAN 5	14.71	39.51	24.03	83.87	31.26	29.13	10.48	33.28	
KAN 7	3.65	30.87	20.56	71.54	9.56	26.32	2.28	23.54	
FCN 3	0.24	27.87	33.99	101.28	5.35	18.66	0.07	26.78	
FCN 5	8.81	35.18	28.74	104.86	18.70	23.96	3.73	32.00	
FCN 7	0.52	26.69	24.54	77.73	6.15	18.60	0.13	22.05	
GRU 3	4.25	26.53	39.34	100.21	11.97	18.97	0.04	28.76	
GRU 5	10.73	36.04	40.65	102.76	14.37	23.50	4.21	33.18	
GRU 7	0.12	24.61	39.38	100.57	18.88	21.75	0.43	29.39	
exp. sm.	200.83	84.13	69.57	73.55	60.55	68.21	37.96	84.97	74.39
exp. (Holt)	19.92	242.55	22.50	47.47	89.88	33.74	30.54	69.51	
ARIMA	181.20	55.34	247.89	78.21	46.13	41.42	14.67	94.98	
r. forest	65.78	47.62	14.83	36.94	109.12	18.18	44.28	48.11	

Table A.30
MAPE values for PmGdaLeczkow air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.23	0.42	0.48	3.97	4.79	20.93	4.02	2.16	–	–
max	1.24	0.56	2.93	25.81	30.30	66.57	15.45	5.22	–	–
TCN 3	13.65	7.02	23.26	21.41	22.94	17.40	21.72	12.58	17.50	9.97
TCN 5	15.11	6.94	17.04	29.67	26.85	22.38	21.73	16.15	19.48	
TCN 7	20.73	7.83	17.09	26.43	21.75	18.18	20.68	11.51	18.03	
KAN 3	11.97	5.75	4.17	2.95	3.90	17.40	15.92	8.28	8.79	
KAN 5	11.28	5.46	8.39	12.89	12.60	18.02	16.34	8.66	11.71	
KAN 7	12.32	5.47	1.39	3.88	5.06	17.34	15.30	8.11	8.61	
FCN 3	0.21	4.18	0.00	2.52	2.89	17.89	2.02	5.91	4.45	
FCN 5	11.45	4.28	5.83	3.50	4.25	18.77	12.43	8.59	8.64	
FCN 7	0.38	3.87	0.50	1.32	3.11	18.15	2.60	7.64	4.70	
GRU 3	0.16	3.68	0.77	1.02	1.10	17.77	11.83	7.33	5.46	
GRU 5	6.86	4.26	3.35	0.74	3.06	18.39	16.29	6.54	7.44	
GRU 7	0.78	5.58	0.90	3.05	1.12	17.68	5.20	4.21	4.81	
exp. sm.	24.55	4.78	26.65	64.75	51.91	27.06	16.22	39.49	31.93	31.95
exp. (Holt)	65.60	2.97	80.82	73.53	71.26	27.79	14.31	42.09	47.30	
ARIMA	17.46	2.27	15.65	40.06	41.84	21.06	8.69	28.95	22.00	
r. forest	24.12	2.37	15.12	42.35	47.63	27.41	11.99	41.72	26.59	

Table A.31
MAPE values for SlCzestoBacz air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.37	0.21	0.72	4.39	5.49	25.00	7.74	2.57	–	–
max	2.38	0.42	1.99	17.56	19.44	53.58	26.66	19.64	–	–
TCN 3	15.90	6.62	16.27	16.18	14.58	14.16	19.68	22.11	15.62	14.96
TCN 5	17.84	12.39	16.49	155.81	13.49	16.54	23.46	33.22	36.15	
TCN 7	11.14	11.61	22.19	18.94	14.60	14.92	27.50	23.99	17.99	
KAN 3	0.13	5.44	6.74	50.01	2.70	13.72	14.17	6.64	12.44	
KAN 5	8.12	9.28	16.39	125.28	8.32	13.48	13.92	14.40	26.15	

(continued on next page)

Table A.31 (*continued*)

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
KAN 7	1.51	5.78	1.89	49.73	3.61	13.15	12.07	4.62	11.54	
FCN 3	0.72	5.21	0.55	20.45	1.78	14.30	6.75	3.92	6.71	
FCN 5	5.99	5.48	9.43	33.68	6.46	14.29	11.54	9.04	11.99	
FCN 7	2.05	5.33	3.31	17.34	1.35	14.18	5.31	1.17	6.26	
GRU 3	0.91	4.92	0.38	33.97	1.97	14.09	6.14	0.35	7.84	
GRU 5	1.69	5.00	6.11	123.01	1.53	14.15	11.06	1.62	20.52	
GRU 7	0.00	5.38	0.09	4.55	0.94	14.18	23.29	5.04	6.68	
exp. sm.	38.88	10.60	16.58	26.18	51.19	23.04	26.74	100.45	38.21	28.83
exp. (Holt)	45.07	6.26	17.29	34.55	68.19	17.27	27.06	20.09	28.75	
ARIMA	24.05	5.15	4.43	17.99	53.59	16.59	20.08	8.59	18.93	
r. forest	38.45	4.46	11.44	31.06	69.53	17.09	25.21	39.95	29.45	

Table A.32
MAPE values for ZpKosz air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.38	0.06	0.45	3.03	4.19	30.63	5.19	0.18	–	–
max	1.50	0.16	13.39	19.92	38.06	74.27	14.58	7.77	–	–
TCN 3	18.54	11.97	22.88	20.92	25.48	15.90	25.62	25.55	20.86	11.53
TCN 5	20.74	15.36	27.28	25.39	31.34	18.38	20.11	28.03	23.33	
TCN 7	16.91	20.23	32.10	25.21	33.07	11.71	16.90	26.86	22.87	
KAN 3	9.31	13.06	8.92	1.83	4.54	16.10	11.23	0.40	8.18	
KAN 5	12.72	13.77	10.52	9.67	10.10	16.42	8.87	22.23	13.04	
KAN 7	9.40	15.70	8.02	0.75	3.63	15.97	4.19	1.06	7.34	
FCN 3	6.55	12.32	6.76	1.35	3.38	16.68	6.01	0.18	6.65	
FCN 5	11.14	13.98	6.36	5.18	9.19	17.17	9.60	1.98	9.32	
FCN 7	2.35	13.03	6.64	0.86	3.52	16.56	1.24	0.10	5.54	
GRU 3	2.18	13.06	7.73	0.31	4.43	16.53	5.68	4.85	6.85	
GRU 5	11.19	13.57	7.97	0.80	8.29	16.90	3.40	0.55	7.83	
GRU 7	7.77	12.39	10.07	0.25	4.10	16.53	1.27	0.30	6.59	
exp. sm.	31.32	13.84	181.02	38.37	123.50	27.90	31.86	216.91	83.09	72.30
exp. (Holt)	28.64	15.42	137.53	75.62	126.37	30.46	42.45	119.14	71.95	
ARIMA	22.70	10.50	169.69	37.27	99.10	26.34	25.38	118.05	63.63	
r. forest	29.33	13.84	116.74	47.93	113.53	23.14	43.88	175.93	70.54	

A.2.3. Full RMSE results

Table A.33
RMSE values for DsJelGorOgin air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.30	0.25	0.00	5.50	6.33	10.87	9.43	3.17	–	–
max	2.86	0.63	9.59	24.85	30.51	54.82	46.26	8.73	–	–
TCN 3	0.30	0.09	0.45	2.24	3.33	4.55	6.81	1.01	2.35	0.10
TCN 5	0.29	0.09	0.94	2.39	3.95	6.17	9.43	1.03	3.04	
TCN 7	0.29	0.07	0.73	2.44	4.69	7.06	6.69	0.99	2.87	
KAN 3	0.01	0.04	0.10	0.44	0.53	3.13	3.29	0.81	1.04	
KAN 5	0.10	0.07	0.20	1.43	3.66	4.65	6.08	0.86	2.13	
KAN 7	0.00	0.04	0.08	0.24	0.54	3.37	3.08	0.82	1.02	
FCN 3	0.00	0.02	0.00	0.38	0.47	4.51	5.58	0.74	1.46	
FCN 5	0.03	0.08	0.17	1.74	1.84	5.33	5.70	0.73	1.95	
FCN 7	0.01	0.02	0.03	0.17	0.59	4.35	5.54	0.75	1.43	
GRU 3	0.00	0.04	0.05	1.20	0.56	5.16	5.48	0.68	1.65	
GRU 5	0.05	0.08	0.19	1.38	2.97	5.37	5.66	0.38	2.01	
GRU 7	0.00	0.04	0.09	0.20	1.06	5.33	5.71	0.25	1.59	
exp. sm.	0.54	0.09	3.14	3.61	6.18	9.92	8.62	0.96	4.13	0.24
exp. (Holt)	0.29	0.05	3.69	5.63	9.13	14.75	11.64	2.16	5.92	
ARIMA	0.19	0.04	4.24	6.36	13.35	14.59	15.29	0.54	6.83	
r. forest	0.31	0.05	0.63	1.93	12.04	3.23	5.52	0.64	3.04	

Table A.34
RMSE values for DsWrocWybCon air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.28	0.22	0.33	6.47	7.64	21.40	9.97	4.56	–	–
max	0.98	0.48	3.58	27.80	31.39	58.36	25.32	7.85	–	–

(*continued on next page*)

Table A.34 (*continued*)

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
TCN 3	0.09	0.06	0.52	3.88	5.20	6.97	3.88	0.78	2.67	0.13
TCN 5	0.10	0.04	0.50	3.31	5.76	8.03	3.97	0.78	2.81	
TCN 7	0.12	0.05	0.48	4.19	5.30	7.34	4.41	0.90	2.85	
KAN 3	0.11	0.03	0.05	1.28	1.48	6.12	0.07	0.51	1.21	
KAN 5	0.15	0.04	0.22	2.46	2.65	6.46	2.22	0.47	1.83	
KAN 7	0.07	0.03	0.05	1.61	1.23	6.45	1.87	0.51	1.48	
FCN 3	0.03	0.03	0.00	1.12	1.32	6.49	0.92	0.49	1.30	
FCN 5	0.11	0.03	0.11	2.32	2.50	6.76	2.19	0.45	1.81	
FCN 7	0.01	0.03	0.01	1.01	1.39	6.48	0.95	0.58	1.31	
GRU 3	0.04	0.02	0.01	0.92	1.40	6.47	1.91	0.53	1.41	
GRU 5	0.02	0.03	0.10	2.03	2.52	6.74	2.18	0.42	1.75	
GRU 7	0.01	0.03	0.01	0.91	2.27	6.48	2.72	0.39	1.60	
exp. sm.	0.16	0.05	1.06	4.88	5.46	6.39	3.50	0.49	2.75	0.20
exp. (Holt)	0.12	0.03	0.35	7.61	7.09	13.10	7.57	0.37	4.53	
ARIMA	0.07	0.02	0.30	6.98	7.20	10.45	7.52	0.16	4.09	
r. forest	0.09	0.02	0.34	3.07	2.70	3.68	1.65	0.87	1.55	

Table A.35
RMSE values for LuGorzKosGdy air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.01	0.20	0.51	14.50	20.65	1.89	7.84	4.20	–	–
max	0.26	0.60	17.33	48.85	69.75	45.00	37.09	8.44	–	–
TCN 3	0.03	0.04	2.27	6.16	12.72	5.91	2.39	1.15	3.84	0.13
TCN 5	0.03	0.05	3.25	9.43	12.98	6.92	2.83	1.08	4.57	
TCN 7	0.04	0.04	1.69	6.81	10.12	7.48	3.69	0.90	3.85	
KAN 3	0.03	0.03	0.01	6.33	9.63	3.94	1.30	0.61	2.74	
KAN 5	0.03	0.04	0.27	6.08	9.72	5.03	1.78	0.70	2.96	
KAN 7	0.03	0.03	0.13	6.26	9.38	4.97	1.25	0.63	2.84	
FCN 3	0.03	0.03	0.00	6.15	11.26	3.62	1.52	0.67	2.91	
FCN 5	0.03	0.04	0.08	7.47	11.72	5.04	2.15	0.65	3.40	
FCN 7	0.03	0.03	0.00	5.94	11.26	3.63	1.59	0.62	2.89	
GRU 3	0.01	0.02	0.00	9.47	11.26	3.58	1.56	0.83	3.34	
GRU 5	0.03	0.04	0.01	9.28	11.71	4.88	1.07	0.64	3.46	
GRU 7	0.03	0.03	0.00	9.48	11.27	3.68	1.51	0.60	3.32	
exp. sm.	0.04	0.07	6.08	11.52	12.31	12.17	12.90	1.55	7.08	0.19
exp. (Holt)	0.03	0.04	6.10	12.03	12.74	3.64	2.29	0.55	4.68	
ARIMA	0.01	0.02	2.39	12.64	7.07	1.17	1.37	0.47	3.14	
r. forest	0.03	0.04	2.62	4.90	8.17	16.41	2.44	0.50	4.39	

Table A.36
RMSE values for LuSulecDudka air quality monitoring station.

	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.42	3.50	5.95	11.20	22.75	8.67	2.77	–	–
max	1.09	19.80	33.66	63.36	30.66	63.24	4.90	–	–
TCN 3	0.05	1.87	2.80	4.36	3.78	3.24	0.42	2.36	0.11
TCN 5	0.06	1.76	3.62	5.43	4.40	3.12	0.47	2.69	
TCN 7	0.05	1.49	3.42	3.88	4.69	3.44	0.47	2.49	
KAN 3	0.04	0.20	0.52	4.69	1.69	3.05	0.29	1.50	
KAN 5	0.04	0.38	0.61	4.51	1.84	3.20	0.31	1.56	
KAN 7	0.03	0.20	0.58	4.73	1.69	3.16	0.23	1.52	
FCN 3	0.03	0.26	0.92	6.55	1.69	3.38	0.27	1.87	
FCN 5	0.04	0.33	0.72	6.56	1.84	3.28	0.31	1.87	
FCN 7	0.03	0.31	0.75	6.06	1.69	3.38	0.23	1.78	
GRU 3	0.03	0.17	0.71	5.73	1.69	3.38	0.23	1.71	
GRU 5	0.03	0.28	0.54	6.07	1.84	3.39	0.31	1.78	
GRU 7	0.03	0.12	0.72	6.47	1.69	3.38	0.28	1.81	
exp. sm.	0.41	11.05	4.62	35.85	1.32	25.61	0.36	11.32	0.27
exp. (Holt)	0.04	9.79	18.04	1.48	5.18	5.09	1.19	5.83	
ARIMA	0.03	8.18	1.20	2.23	0.29	2.47	0.88	2.18	
r. forest	0.02	7.57	1.36	2.48	0.77	2.76	0.28	2.18	

Table A.37
RMSE values for MpZakopaSien air quality monitoring station.

	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.24	0.85	6.32	7.70	1.83	5.26	0.51	–	–
max	1.66	124.44	58.62	248.70	62.12	102.34	14.86	–	–

(*continued on next page*)

Table A.37 (*continued*)

	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
TCN 3	0.19	22.34	11.17	40.45	9.11	15.95	2.59	14.54	0.17
TCN 5	0.26	25.62	13.01	51.24	9.23	18.95	2.09	17.20	
TCN 7	0.29	19.19	13.58	36.04	10.53	17.33	2.47	14.20	
KAN 3	0.03	23.08	9.62	70.56	4.86	19.26	0.76	18.31	
KAN 5	0.14	23.98	9.57	72.15	8.10	19.99	0.88	19.26	
KAN 7	0.04	22.95	9.62	69.66	5.09	19.42	0.32	18.16	
FCN 3	0.00	23.76	13.20	76.74	5.58	19.51	0.02	19.83	
FCN 5	0.07	22.76	11.47	77.32	8.85	20.35	0.55	20.20	
FCN 7	0.01	23.87	10.91	71.04	5.43	19.48	0.02	18.68	
GRU 3	0.04	23.79	15.62	76.48	8.90	19.93	0.02	20.68	
GRU 5	0.12	25.72	15.71	76.82	9.00	20.69	1.24	21.33	
GRU 7	0.00	23.67	15.62	76.57	7.31	20.30	0.05	20.50	
exp. sm.	0.63	27.88	18.82	66.85	14.50	26.14	2.67	22.50	0.23
exp. (Holt)	0.22	48.74	8.59	37.79	9.41	15.02	2.42	17.46	
ARIMA	0.62	12.51	16.50	24.05	10.45	11.01	1.38	10.93	
r. forest	0.37	17.44	5.53	62.05	16.73	8.41	8.25	16.97	

Table A.38
RMSE values for PmGdaLeczkow air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.23	0.42	0.48	3.97	4.79	20.93	4.02	2.16	–	–
max	1.24	0.56	2.93	25.81	30.30	66.57	15.45	5.22	–	–
TCN 3	0.11	0.04	0.29	2.78	3.17	8.20	2.62	0.44	2.20	0.13
TCN 5	0.11	0.04	0.23	3.82	4.71	11.58	2.94	0.67	3.01	
TCN 7	0.14	0.05	0.24	3.54	3.48	8.33	2.53	0.43	2.34	
KAN 3	0.09	0.03	0.06	1.00	1.65	10.26	2.15	0.31	1.94	
KAN 5	0.09	0.03	0.14	2.27	2.92	10.32	2.27	0.33	2.30	
KAN 7	0.09	0.03	0.02	1.76	1.61	10.23	2.05	0.30	2.01	
FCN 3	0.00	0.03	0.00	1.11	1.66	10.43	0.69	0.21	1.77	
FCN 5	0.09	0.03	0.10	0.97	1.48	10.59	1.88	0.33	1.93	
FCN 7	0.01	0.02	0.01	0.89	1.80	10.51	0.74	0.29	1.78	
GRU 3	0.00	0.02	0.01	0.72	1.10	10.40	1.69	0.28	1.78	
GRU 5	0.08	0.03	0.05	0.56	2.31	10.47	2.15	0.24	1.98	
GRU 7	0.01	0.03	0.01	1.98	1.02	10.37	1.21	0.17	1.85	
exp. sm.	0.20	0.03	0.41	5.87	6.09	12.59	2.21	0.90	3.54	0.21
exp. (Holt)	0.18	0.02	0.36	6.14	4.72	12.77	1.76	1.02	3.37	
ARIMA	0.11	0.02	0.33	3.73	6.05	14.39	1.17	0.82	3.33	
r. forest	0.18	0.02	0.36	4.39	6.00	13.18	1.73	0.99	3.36	

Table A.39
RMSE values for SlCzestoBacz air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.37	0.21	0.72	4.39	5.49	25.00	7.74	2.57	–	–
max	2.38	0.42	1.99	17.56	19.44	53.58	26.66	19.64	–	–
TCN 3	0.16	0.03	0.26	21.76	2.26	6.98	4.38	2.73	4.82	0.31
TCN 5	0.15	0.05	0.21	25.90	2.09	7.69	4.58	3.76	5.55	
TCN 7	0.16	0.04	0.28	28.44	2.07	6.51	5.21	3.02	5.72	
KAN 3	0.00	0.02	0.10	26.91	0.48	6.98	2.83	0.80	4.77	
KAN 5	0.08	0.04	0.24	27.64	1.28	6.77	2.65	1.26	5.00	
KAN 7	0.01	0.03	0.03	23.80	0.59	6.64	2.63	0.68	4.30	
FCN 3	0.01	0.02	0.01	11.68	0.38	7.13	1.72	0.55	2.69	
FCN 5	0.06	0.02	0.14	29.43	1.02	7.07	2.32	0.81	5.11	
FCN 7	0.03	0.02	0.05	21.20	0.34	7.10	1.77	0.17	3.83	
GRU 3	0.01	0.02	0.01	12.75	0.39	7.07	1.90	0.07	2.78	
GRU 5	0.02	0.02	0.09	26.04	0.32	7.04	2.37	0.16	4.51	
GRU 7	0.00	0.02	0.00	5.08	0.31	7.10	4.14	0.63	2.16	
exp. sm.	0.56	0.04	0.26	20.55	4.31	13.11	6.61	4.92	6.29	0.50
exp. (Holt)	0.74	0.02	0.21	25.82	5.66	12.09	7.51	2.56	6.83	
ARIMA	0.50	0.02	0.06	28.62	5.21	12.06	7.33	1.37	6.89	
r. forest	0.62	0.02	0.18	31.02	5.78	11.68	7.20	6.83	7.92	

Table A.40
RMSE values for ZpKosz air quality monitoring station.

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean(1)	mean(2)
min	0.38	0.06	0.45	3.03	4.19	30.63	5.19	0.18	–	–
max	1.50	0.16	13.39	19.92	38.06	74.27	14.58	7.77	–	–

(*continued on next page*)

Table A.40 (*continued*)

	C_6H_6	CO	NO	NO_2	NO_x	O_3	PM_{10}	SO_2	mean$^{(1)}$	mean$^{(2)}$
TCN 3	0.14	0.02	1.23	2.02	5.69	7.14	2.94	1.03	2.53	0.11
TCN 5	0.15	0.02	1.39	2.84	5.25	9.20	2.40	1.00	2.78	
TCN 7	0.12	0.03	1.60	2.73	5.80	5.85	2.06	0.99	2.40	
KAN 3	0.08	0.02	0.20	0.20	3.05	9.84	1.60	0.05	1.88	
KAN 5	0.11	0.02	0.24	0.91	3.58	9.86	1.19	0.59	2.06	
KAN 7	0.08	0.02	0.17	0.21	1.95	9.81	0.53	0.02	1.60	
FCN 3	0.04	0.02	0.16	0.27	2.32	10.01	1.25	0.01	1.76	
FCN 5	0.10	0.02	0.14	0.57	3.75	10.10	1.17	0.10	1.99	
FCN 7	0.02	0.02	0.16	0.21	2.38	9.97	0.28	0.01	1.63	
GRU 3	0.02	0.02	0.17	0.23	3.22	9.96	1.22	0.17	1.88	
GRU 5	0.10	0.02	0.16	0.40	3.61	10.00	0.51	0.04	1.85	
GRU 7	0.07	0.02	0.20	0.19	3.26	9.96	0.17	0.03	1.74	
exp. sm.	0.39	0.02	3.65	3.54	12.10	11.56	5.21	3.52	5.00	0.33
exp. (Holt)	0.51	0.02	2.83	4.38	12.02	14.58	3.88	2.53	5.09	
ARIMA	0.39	0.01	3.76	3.71	11.57	13.79	5.46	2.66	5.17	
r. forest	0.51	0.02	2.88	4.50	11.92	10.47	4.02	3.11	4.68	

References

Abirami, S., Chitra, P., 2021. Regional air quality forecasting using spatiotemporal deep learning. J. Clean. Prod. 283, 125341.

Ahmadi, M.H., Jashnani, H., Chau, K.-W., Kumar, R., Rosen, M.A., 2023. Carbon dioxide emissions prediction of five middle eastern countries using artificial neural networks. Energy Sources Part A 45, 9513–9525.

Bai, S., Kolter, J.Z., Koltun, V., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling arXiv preprint arXiv:1803.01271.

Baklanov, A., Zhang, Y., 2020. Advances in air quality modeling and forecasting. Glob. Trans. 2, 261–270.

Bekkar, A., Hssina, B., Douzi, S., Douzi, K., 2021. Air-pollution prediction in smart city, deep learning approach. J. Big Data 8, 1–21.

Bhatti, U.A., Yan, Y., Zhou, M., Ali, S., Hussain, A., Qingsong, H., Yu, Z., Yuan, L., 2021. Time series analysis and forecasting of air pollution particulate matter (pm 2.5): an sarima and factor analysis approach. Ieee Access 9, 41019–41031.

Chang, Y.-S., Chiao, H.-T., Abimannan, S., Huang, Y.-P., Tsai, Y.-T., Lin, K.-M., 2020. An lstm-based aggregated model for air pollution forecasting. Atmos. Pollut. Res. 11, 1451–1463.

Cho, J.H., Moon, J.W., 2022. Integrated artificial neural network prediction model of indoor environmental quality in a school building. J. Clean. Prod. 344, 131083.

Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. URL. `https://arxiv.org/abs/1412.3555`. `arXiv:1412.3555`.

Czernecki, B., Marosz, M., Jędruszkiewicz, J., et al., 2021. Assessment of machine learning algorithms in short-term forecasting of pm10 and pm2. 5 concentrations in selected polish agglomerations. Aerosol Air Qual. Res. 21, 200586.

Dobrea, M., Bădicu, A., Barbu, M., Subea, O., Bălănescu, M., Suciu, G., Bîrdici, A., Orza, O., Dobre, C., 2020. Machine learning algorithms for air pollutants forecasting. In: 2020 IEEE 26th International Symposium for Design and Technology in Electronic Packaging (SIITME). IEEE, pp. 109–113.

Du, S., Li, T., Yang, Y., Horng, S.-J., 2019. Deep air quality forecasting using hybrid deep learning framework. IEEE Trans. Knowl. Data Eng. 33, 2412–2424.

Gaj, K., Miller, U., Sówka, I., 2020. Progressing Climate Changes and Deteriorating Air Quality as One of the Biggest Challenges of Sustainable Development of the Present Decade in Poland.

Gardner Jr., E.S., 1985. Exponential smoothing: The state of the art. J. Forecast. 4, 1–28. https://doi.org/10.1002/for.3980040103.

Gilik, A., Ogrenci, A.S., Ozmen, A., 2022. Air quality prediction using cnn+ lstm-based hybrid deep learning architecture. Environ. Sci. Pollut. Res. 1–19.

Gu, Y., Li, B., Meng, Q., 2022. Hybrid interpretable predictive machine learning model for air pollution prediction. Neurocomputing 468, 123–136.

Harishkumar, K., Yogesh, K., Gad, I., et al., 2020. Forecasting air pollution particulate matter (pm2. 5) using machine learning regression models. Procedia Comput. Sci. 171, 2057–2066.

Heuvelmans, M.A., van Ooijen, P.M., Ather, S., Silva, C.F., Han, D., Heussel, C.P., Hickes, W., Kauczor, H.-U., Novotny, P., Peschl, H., et al., 2021. Lung cancer prediction by deep learning to identify benign lung nodules. Lung Cancer 154, 1–4.

Heydari, A., Majidi Nezhad, M., Astiaso Garcia, D., Keynia, F., De Santoli, L., 2022. Air pollution forecasting application based on deep learning model and optimization algorithm. Clean Techn. Environ. Policy 1–15.

Hong Zhang, P.W.Y.Q., Zhang, Sheng, Wang, H., 2017. Forecasting of particulate matter time series using wavelet analysis and wavelet-arma/arima model in taiyuan, china. J. Air Waste Manage. Assoc. 67, 776–788. https://doi.org/10.1080/10962247.2017.1292968.

Ke, H., Gong, S., He, J., Zhang, L., Cui, B., Wang, Y., Mo, J., Zhou, Y., Zhang, H., 2022. Development and application of an automated air quality forecasting system based on machine learning. Sci. Total Environ. 806, 151204.

Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S., 2017. Self-normalizing neural networks. CoRR, abs/1706.02515. URL. `http://arxiv.org/abs/1706.02515`. `arXiv:1706.02515`.

Kruskal, W.H., 1952. A nonparametric test for the several sample problem. Ann. Math. Stat. 23, 525–540. https://doi.org/10.1214/aoms/1177729332.

Kruskal, W., Wallis, W., 1952. Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 583–621.

Kryza, M., Werner, M., Guzikowski, J., 2020. Assimilation of meteorological data in online integrated atmospheric transport model—Example of air quality forecasts for Poland. In: Air Pollution Modeling and its Application XXVI, vol. 36. Springer, pp. 273–278.

Kujawska, J., Kulisz, M., Oleszczuk, P., Cel, W., 2022. Machine learning methods to forecast the concentration of pm10 in Lublin, Poland. Energies 15, 6428.

Kurnaz, G., Demir, A.S., 2022. Prediction of so2 and pm10 air pollutants using a deep learning-based recurrent neural network: case of industrial city Sakarya. Urban Clim. 41, 101051.

Lea, C., Vidal, R., Reiter, A., Hager, G.D., 2016. Temporal convolutional networks: A unified approach to action segmentation. In: Hua, G., Jégou, H. (Eds.), Computer Vision - ECCV 2016 Workshops - Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III (pp. 47–54). volume 9915 of Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-319-49409-8_7.

Lea, C., Flynn, M.D., Vidal, R., Reiter, A., Hager, G.D., 2017. Temporal convolutional networks for action segmentation and detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017. IEEE Computer Society, pp. 1003–1012. https://doi.org/10.1109/CVPR.2017.113.

Lee, D., Kang, S., Shin, J., 2017. Using deep learning techniques to forecast environmental consumption level. Sustainability 9, 1894.

Lee, M., Lin, L., Chen, C.-Y., Tsao, Y., Yao, T.-H., Fei, M.-H., Fang, S.-H., 2020. Forecasting air quality in Taiwan by using machine learning. Sci. Rep. 10, 4153.

Liao, Q., Zhu, M., Wu, L., Pan, X., Tang, X., Wang, Z., 2020. Deep learning for air quality forecasts: a review. Curr. Pollut. Rep. 6, 399–409.

Lima, S., Gonçalves, A.M., Costa, M., 2019. Time series forecasting using Holt-Winters exponential smoothing: an application to economic data. AIP Conf. Proc. 2186, 090003. https://doi.org/10.1063/1.5137999.
Liu, H., Yin, S., Chen, C., Duan, Z., 2020. Data multi-scale decomposition strategies for air pollution forecasting: a comprehensive review. J. Clean. Prod. 277, 124023.
Liu, H., Yan, G., Duan, Z., Chen, C., 2021. Intelligent modeling strategies for forecasting air quality time series: a review. Appl. Soft Comput. 102, 106957.
Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T.Y., Tegmark, M., 2024. Kan: Kolmogorov-Arnold Networks. URL. https://arxiv.org/abs/2404.19756. arXiv:2404.19756.
Majewski, G., Szeląg, B., Mach, T., Rogula-Kozłowska, W., Anioł, E., Bihałowicz, J., Dmochowska, A., Bihałowicz, J.S., 2021. Predicting the number of days with visibility in a specific range in Warsaw (Poland) based on meteorological and air quality data. Front. Environ. Sci. 9, 623094.
Mao, W., Jiao, L., Wang, W., Wang, J., Tong, X., Zhao, S., 2021. A hybrid integrated deep learning model for predicting various air pollutants. GISci. Remote Sens. 58, 1395–1412.
Masood, A., Ahmad, K., 2021. A review on emerging artificial intelligence (ai) techniques for air pollution forecasting: fundamentals, application and performance. J. Clean. Prod. 322, 129072.
Méndez, M., Merayo, M.G., Núñez, M., 2023. Machine learning algorithms to forecast air quality: a survey. Artif. Intell. Rev. 56, 10031–10066.
Mishra, D., Goyal, P., 2015. Development of artificial intelligence based no2 forecasting models at taj mahal, agra. Atmos. Pollut. Res. 6, 99–106.
Murillo-Escobar, J., Sepulveda-Suescun, J., Correa, M., Orrego-Metaute, D., 2019. Forecasting concentrations of air pollutants using support vector regression improved with particle swarm optimization: case study in aburrá valley, Colombia. Urban Clim. 29, 100473.
Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In: Fürnkranz, J., Joachims, T. (Eds.), Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21–24, 2010, Haifa, Israel. Omnipress, pp. 807–814. URL. https://icml.cc/Conferences/2010/papers/432.pdf.
Nilashi, M., Ahmadi, H., Manaf, A.A., Rashid, T.A., Samad, S., Shahmoradi, L., Aljojo, N., Akbari, E., 2020. Coronary heart disease diagnosis through self-organizing map and fuzzy support vector machine with incremental updates. Int. J. Fuzzy Syst. 22, 1376–1388.
Pak, U., Ma, J., Ryu, U., Ryom, K., Juhyok, U., Pak, K., Pak, C., 2020. Deep learning-based pm2. 5 prediction considering the spatiotemporal correlations: a case study of Beijing, China. Sci. Total Environ. 699, 133561.
Połednik, B., 2022. Emissions of air pollution in industrial and rural region in Poland and health impacts. J. Ecol. Eng. 23, 250–258.
Qi, Y., Li, Q., Karimian, H., Liu, D., 2019. A hybrid model for spatiotemporal forecasting of pm2. 5 based on graph convolutional neural network and long short-term memory. Sci. Total Environ. 664, 1–10.
Rigatti, S.J., 2017. Random Forest. J. Insur. Med. 47, 31–39. https://doi.org/10.17849/insm-47-01-31-39.1.
Saxena, N., 2021. Efficient downscaling of satellite oceanographic data with convolutional neural networks. SIGSPATIAL Spec. 12, 46–47.
Shams, S.R., Jahani, A., Kalantary, S., Moeinaddini, M., Khorasani, N., 2021. The evaluation on artificial neural networks (ann) and multiple linear regressions (mlr) models for predicting so2 concentration. Urban Clim. 37, 100837.
Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52, 591–611.
Sharma, M., Jain, S., Mittal, S., Sheikh, T.H., 2021. Forecasting and prediction of air pollutants concentrates using machine learning techniques: The case of India. In: IOP Conference Series: Materials Science and Engineering, vol. 1022. IOP Publishing, p. 012123.
Shi, K., Wu, L., 2020. Forecasting air quality considering the socio-economic development in Xingtai. Sustain. Cities Soc. 61, 102337.
Shumway, R.H., Stoffer, D.S., 2017. Arima models. In: Time Series Analysis and its Applications: With R Examples. Springer International Publishing, Cham, pp. 75–163. https://doi.org/10.1007/978-3-319-52452-8_3.
Sohn, S.H., Oh, S.C., Yeo, Y.-K., 1999. Prediction of air pollutants by using an artificial neural network. Korean J. Chem. Eng. 16, 382–387.
Song, C., Fu, X., 2020. Research on different weight combination in air quality forecasting models. J. Clean. Prod. 261, 121169.
Subramaniam, S., Raju, N., Ganesan, A., Rajavel, N., Chenniappan, M., Prakash, C., Pramanik, A., Basak, A.K., Dixit, S., 2022. Artificial intelligence technologies for forecasting air pollution and human health: a narrative review. Sustainability 14. https://doi.org/10.3390/su14169951.
Usmani, R.S.A., Pillai, T.R., Hashem, I.A.T., Marjani, M., Shaharudin, R., Latif, M.T., 2021. Air pollution and cardiorespiratory hospitalization, predictive modeling, and analysis using artificial intelligence techniques. Environ. Sci. Pollut. Res. 28, 56759–56771.
Wang, J., Niu, T., Wang, R., 2017. Research and application of an air quality early warning system based on a modified least squares support vector machine and a cloud model. Int. J. Environ. Res. Public Health 14, 249.
Yang, H., Zhu, Z., Li, C., Li, R., 2020. A novel combined forecasting system for air pollutants concentration based on fuzzy theory and optimization of aggregation weight. Appl. Soft Comput. 87, 105972.
Zeinalnezhad, M., Chofreh, A.G., Goni, F.A., Klemeš, J.J., 2020. Air pollution prediction using semi-experimental regression model and adaptive neuro-fuzzy inference system. J. Clean. Prod. 261, 121218.
Zhang, L., Lin, J., Qiu, R., Hu, X., Zhang, H., Chen, Q., Tan, H., Lin, D., Wang, J., 2018. Trend analysis and forecast of pm2. 5 in Fuzhou, China using the Arima model. Ecol. Indic. 95, 702–710.
Zhu, D., Cai, C., Yang, T., Zhou, X., 2018. A machine learning approach for air quality prediction: model regularization and optimization. Big Data Cognit. Comput. 2, 5.

A.3 Artykuł A3 – Załącznik 3

Environmental Science and Pollution Research
https://doi.org/10.1007/s11356-025-36265-y

RESEARCH ARTICLE

Forecasting the concentration of the components of the particulate matter in Poland using neural networks

Jarosław Bernacki[1]

Received: 28 October 2024 / Accepted: 9 March 2025

Abstract
Air pollution is a significant global challenge with profound impacts on human health and the environment. Elevated concentrations of various air pollutants contribute to numerous premature deaths each year. In Europe, and particularly in Poland, air quality remains a critical concern due to pollutants such as particulate matter (PM), which pose serious risks to public health and ecological systems. Effective control of PM emissions and accurate forecasting of their concentrations are essential for improving air quality and supporting public health interventions. This paper presents four advanced deep learning-based forecasting methods: extended long short-term memory network (xLSTM), Kolmogorov-Arnold network (KAN), temporal convolutional network (TCN), and variational autoencoder (VAE). Using data from eight cities in Poland, we evaluate our methods' ability to predict particulate matter concentrations through extensive experiments, utilizing statistical hypothesis testing and error metrics such as mean absolute error (MAE) and root mean square error (RMSE). Our findings demonstrate that these methods achieve high prediction accuracy, significantly outperforming several state-of-the-art algorithms. The proposed forecasting framework offers practical applications for policymakers and public health officials by enabling timely interventions to decrease pollution impacts and enhance urban air quality management.

Keywords Air pollution · Particulate matter · Forecasting · Predicting · Machine learning · Deep models

Introduction

Air pollution significantly impacts human health, with high concentrations of various pollutants, especially particulate matter (PM_x), being responsible for a substantial number of deaths each year. This issue is global in scope, and in Europe, air quality is notably poor in many regions, especially urban areas (Połednik 2022). In numerous countries, the detrimental effects of poor air quality are often overlooked, partly due to low public awareness and understanding of the problem. A contributing factor to this may be insufficient knowledge about the severity of air pollution. The processes of industrialization and automobile transportation, particularly those involving fuel combustion, contribute to the rise in many air pollutants, including benzene (C_6H_6), carbon monoxide (CO), nitrogen oxides (NO, NO_2, NO_x), ozone (O_3), particulate matter 10 (PM_{10}), sulfur dioxide (SO_2), and others. As a result, air pollution is closely monitored and analyzed at air quality monitoring stations, where the air quality is assessed based on data collected from these stations.

Particulate matter is fine solid particles or liquid droplets suspended in the air. It is a serious environmental problem, especially in urban and industrial areas. Particulate matter is divided into several categories depending on the size of the particles: particulate matter 10 (PM_{10})—aerodynamic diameter of less than 10 micrometers; particulate matter 2.5 ($PM_{2.5}$)—particles with an aerodynamic diameter of less than 2.5 micrometers; particulate matter 1 (PM_1)—particles with an aerodynamic diameter of less than 1 micrometer.

In Poland, the problem of air pollution with particulate matter (PM) is serious, especially in the winter season. This is mainly due to the so-called low emissions, i.e., emissions from domestic heating furnaces burning coal and wood, often of low quality. Other sources are exhaust fumes from combustion engines, especially in large cities, and emissions from industrial plants or power plants. In addition, there are emissions from road transport and industry. Poland is one of

Responsible editor: Marcus Schulz

✉ Jarosław Bernacki
jaroslaw.bernacki@pcz.pl

[1] Department of Artificial Intelligence, Częstochowa University of Technology, al. Armii Krajowej 36, Częstochowa 42-200, Poland

Published online: 21 March 2025

the European Union countries where the concentrations of particulate matter are among the highest. The greatest problems concern cities in southern and central Poland, such as Kraków, Katowice, Wrocław, and Warsaw. In some of these cities, air quality standards set by the European Union and the World Health Organization (WHO) are regularly exceeded.

Poland is taking various actions to improve air quality, including programs to replace old furnaces with more ecological ones and introduce clean transport zones in cities. However, the effectiveness of these actions is limited and requires further efforts at both the national and local levels. The $PM_{2.5}$ and PM_{10} concentrations are monitored on an ongoing basis by the Chief Inspectorate of Environmental Protection (GIOS),[1] which provides data online. Despite some improvement in recent years, the problem of air pollution in Poland remains a significant challenge.

Therefore, monitoring air quality is a crucial task. One challenge in this area is ensuring the completeness of the monitoring data. Air quality monitoring stations frequently encounter technical difficulties, leading to gaps in measurements. To address this issue, one approach is to approximate the missing values, which is closely related to the task of forecasting. Numerous studies in the literature focus on forecasting air pollutant concentrations (Liu et al. 2021, 2020; Yang et al. 2020), employing a wide range of methods, including machine learning (Dobrea et al. 2020; Sharma et al. 2021), deep learning (Abirami and Chitra 2021; Heydari et al. 2022), and hybrid approaches (Chang et al. 2020; Mao et al. 2021). However, in some cases, the estimation results may not be satisfactory.

In this paper, we propose methods for forecasting the concentration of the particulate matter 10 (PM_{10}) and its selected components in Poland. The analyzed PM components include arsenic (As), cadmium (Cd), nickel (Ni), and lead (Pb). These pollutants were analyzed in the following eight air quality monitoring stations in Poland, located in different Polish cities: Wałbrzych, Toruń, Żagań, Kraków, Mielec, Katowice, Puszcza Borecka, and Szczecin. We have proposed methods utilizing novel approaches for forecasting: extended long short-term memory network (xLSTM), Kolmogorov-Arnold network (KAN), temporal convolutional network (TCN), and variational autoencoder (VAE). Our objective was to generate forecasts for the concentrations of the specified air pollutants in the aforementioned cities. This was achieved by training each network using real data collected from GIOS to predict these concentrations. We then compared the forecasting results with the actual data to assess the quality of the predictions. Each proposed network was tested with three different combinations of input parameters, resulting in over 400 prediction models, which were subsequently analyzed. The evaluation of these models was conducted in two ways: through statistical hypothesis testing and by analyzing error metrics such as mean absolute error (MAE) and root mean square error (RMSE). The statistical hypothesis testing confirmed that there were no significant differences between the original data and the forecasts produced by the proposed methods, demonstrating their reliability.

We have also implemented several well-known state-of-the-art algorithms for forecasting, including exponential smoothing, Holt-Winters exponential smoothing, autoregressive integrated moving average (ARIMA), and random forest. Our experimental evaluation demonstrated that the proposed methods produce more accurate forecasts. Statistical hypothesis testing revealed that in some instances, the forecasts generated by state-of-the-art algorithms showed significant statistical differences when compared to the original data. Additionally, the analysis of MAE and RMSE indicated that the proposed methods achieved lower error values compared to the existing algorithms in the literature. These findings suggest that the proposed methods generate more precise forecasts. Such an analysis could be particularly useful for estimating missing data, especially in cases where technical issues at air monitoring stations result in data gaps.

Contribution

This paper introduces four deep learning-based methods for forecasting concentrations of specific particulate matter components in Poland: xLSTM, KAN, TCN, and VAE. Using real-world data from GIOS across eight Polish cities, we evaluated these methods through statistical hypothesis testing and error metrics. The results demonstrate that our proposed approaches achieve high accuracy, offering a significant advancement over existing forecasting techniques for air pollutant concentrations.

Accurate forecasting of PM concentrations is crucial for decreasing air pollution's adverse effects on public health and the environment. While numerous studies have explored PM forecasting, many rely on traditional statistical methods or machine learning approaches that struggle to capture the complex temporal and spatial dependencies inherent in pollution data. This paper addresses this gap by proposing advanced deep learning-based methods specifically designed to model these dependencies, offering improved accuracy and reliability in predicting PM concentrations. Our study focuses on Poland, a country with persistent air quality challenges, using data from eight cities to evaluate the effectiveness of the proposed methods.

Organization of the paper

The paper is organized as follows: The next section describes previous and related works. Section "Proposed methods"

[1] https://www.gov.pl/web/gios-en

presents proposed deep learning methods for forecasting. In Section "Experimental evaluation," the experimental evaluation of the proposed methods is performed. Finally, the last section concludes this work. At the end of the paper, there is an Appendix that stores the large number of evaluation tables.

Previous and related work

A comprehensive survey of various artificial intelligence (AI) techniques for air pollution forecasting is presented in Subramaniam et al. (2022). The authors review a range of methods, including traditional statistical models, such as ARIMA and regression models (Zhang et al. 2018; Masood and Ahmad 2021; Shams et al. 2021; Zeinalnezhad et al. 2020); artificial neural networks which are explored in several studies (Ahmadi et al. 2023; Cho and Moon 2022; Mishra and Goyal 2015; Shams et al. 2021; Sohn et al. 1999; Usmani et al. 2021); support vector machines (Murillo-Escobar et al. 2019; Nilashi et al. 2020; Wang et al. 2017); fuzzy logic models (Masood and Ahmad 2021); deep learning models including various architectures (Bekkar et al. 2021; Du et al. 2019; Gilik et al. 2022; Heuvelmans et al. 2021; Kurnaz and Demir 2022; Lee et al. 2017; Pak et al. 2020; Saxena 2021); and hybrid models combining different approaches (Gu et al. 2022; Qi et al. 2019; Zhu et al. 2018). These models are primarily employed for forecasting major pollutants such as $PM_{2.5}$, PM_{10}, O_3, CO, SO_2, NO_2, and CO_2. Performance evaluation of these models often involves error metrics such as R-squared (the coefficient of determination, R^2), RMSE, MAE, and mean absolute percentage error (MAPE). Additionally, a review of hybrid methods applicable to air quality forecasting is provided in Baklanov and Zhang (2020); Liao et al. (2020); Liu et al. (2021); Méndez et al. (2023).

In Qi et al. (2019), a hybrid model combines neural networks to predict $PM_{2.5}$ concentrations in a given time and space. More precisely, the proposed integrates the graph convolutional networks and long short-term memory networks (GC-LSTM) in order to predict the spatiotemporal variation of $PM_{2.5}$ concentrations. The approach was compared with several state-of-the-art methods, and the results revealed that the GC-LSTM model outperforms other methods in terms of several machine learning measures, including the correlation coefficient R^2. In Liang et al. (2023), the neural networks with meteorological data to predict hourly $PM_{2.5}$ concentrations were used. The experiments were held based on the data coming from Beijing Aotizhongxin. The R^2 measure of the proposed method obtained better values than other approaches. In Yu et al. (2022), an ensemble learning approach to predict $PM_{2.5}$ concentrations, taking into account climate data and air quality indicators from neighboring cities, was presented. The forecasting experiments were held on the data representing $PM_{2.5}$ concentrations in Beijing, Wuhan, and Shenzhen.

In Usha Ruby et al. (2024), a hybrid methodology for forecasting $PM_{2.5}$ concentrations was described. The method uses a gradient-boosted regression tree with convolutional neural networks and fuzzy K-nearest neighbor. The forecasting results were compared with methods such as multiple linear regression, stacked long short-term memory, bidirectional gated recurrent unit, and gradient-boosted regression tree. Results evaluated with the measures including RMSE, MAE, and MAPE confirmed higher prediction accuracy over other methods. In Liu et al. (2019), multi-stage forecasting of $PM_{2.5}$ concentrations using a new hybrid model that combines different machine learning techniques is described. The proposed method uses a wavelet packet decomposition (WPD), particle swarm optimization (PSO), back propagation neural network, and AdaBoost algorithm. Results indicated that the PSO and AdaBoost can effectively enhance the forecasting accuracy.

In Kryza et al. (2021), the forecasting of hourly $PM_{2.5}$ concentrations in southwestern Poland using the uEMEP and EMEP4PL models is considered. The results are compared with data obtained from measurement networks, which allows for the assessment of the forecast accuracy. EMEP4PL is an application of the EMEP MSC-W chemistry transport model specifically tailored for Poland. This model operates using two nested domains: the parent domain, which covers Europe with a grid resolution of $12\,\text{km} \times 12\,\text{km}$, and the nested domain, which is focused on Poland with a finer grid resolution of $4\,\text{km} \times 4\,\text{km}$. The uEMEP model, based on Gaussian modeling principles, enables the downscaling of EMEP4PL results to an even higher spatial resolution using the local fraction approach. The results indicated that the uEMEP forecasts have a smaller bias and higher index of agreement, compared to EMEP. The application of the random forest algorithm to improve the estimates of daily $PM_{2.5}$ concentrations in Poland generated by the EMEP4PL model is described in Vovk et al. (2024).

In Krajny et al. (2024), the short-term forecasting of PM_{10} concentrations in Krakow using sodar Doppler technology, paying attention to the impact of meteorological conditions on air pollution levels, was presented. The discussed method was based on the use of the spectrum, i.e., the set of amplitudes of signals returning to the sodar receiver from the reflection of a single-frequency sound transmission and its characteristic properties depending on the physical state of the atmosphere. The data mining methods were used to generate the forecasts of high concentrations of PM_{10}. In Chae et al. (2021), convolutional neural network models were used to predict real-time PM_{10} and $PM_{2.5}$ concentrations in Poland. The evaluation using the R^2 and RMSE measures confirmed the high detection accuracy. In Udristioiu et al. (2023), there was proposed the use of hybrid models to forecast daily concentrations of various PM fractions and the air quality index (AQI) in Romania. The utilized methods included input variable selection (IVS), machine learning (ML), and regression models. Evaluation using well-known measures such as R^2 or RMSE confirmed the satisfactory results of forecasting the PM concentrations.

In Hong Zhang and Wang (2017), the concentrations of PM_{10} in Taiyuan, China, were analyzed using a wavelet-based ARMA/ARIMA model. The model's performance was evaluated using relative error (RE), RMSE, and normalized root mean square error (NRMSE). The study found that the values of these error metrics were minimal, indicating a high accuracy of the model. In Doreswamy et al. (2020), a method for predicting $PM_{2.5}$ concentrations was proposed using various regression models, including random forest, gradient boosting regression, decision tree regression, and multilayer perceptron regression. The data used for this study were collected from Pingtung station and Newport City in Taiwan. The evaluation of the prediction performance was conducted using MAE and RMSE. An aggregated long short-term memory (LSTM) model for air pollution forecasting was considered in Chang et al. (2020). This model was compared against three popular machine learning methods, with the evaluation based on MAE, RMSE, and MAPE. The research utilized data collected from various locations in Taiwan. In Lee et al. (2020), a gradient-boosting-based machine learning approach for predicting $PM_{2.5}$ concentrations in Taiwan was proposed. The results were also compared with data from Taipei and London. The findings showed similar prediction results for Taipei and London, attributed to the comparable topography of the cities. The evaluation of the results was performed using the coefficient of determination (R^2) and RMSE.

In Bhatti et al. (2021), a method for predicting $PM_{2.5}$ concentrations using seasonal autoregressive integrated moving average (SARIMA) is presented. The effectiveness of the SARIMA model is evaluated through statistical analysis, which demonstrates its efficacy in handling seasonal variations in pollutant concentrations. In Song and Fu (2020), a radial basis neural network (RBNN) is employed for air quality prediction in Zhengzhou and Shanghai, China. The study explores the network's ability to model and forecast air quality based on the collected data from these cities. The method for predicting a comprehensive air quality index in Xingtai, China, is proposed in Shi and Wu (2020). This approach uses the speed of socio-economic development into the forecasting model. The results indicate that the grey multivariable model used in this study provides high forecasting performance. In Ke et al. (2022), a method for daily forecasting of $PM_{2.5}$, PM_{10}, SO_2, NO_2, O_3, and CO based on data from multiple Chinese cities (Beijing, Shanghai, Guangzhou, Chengdu, Xi'an, Wuhan, and Changchun) is described. The study utilizes five machine learning models to generate 3-day pollution level forecasts, and the evaluation confirms that the predictions are satisfactory.

In Sówka et al. (2019), an analysis of particulate matter concentrations in selected Polish cities, including Wrocław and Poznań, from 2014 to 2016, is presented. The study found that the highest concentrations of $PM_{1.0}$, $PM_{2.5}$, and PM_{10} were observed in 2016 in both cities. In Czernecki et al. (2021), machine learning methods for short-term forecasting of $PM_{2.5}$ and PM_{10} in major Polish cities are examined. The study employs AIC-based stepwise regression, tree-based algorithms (random forests and XGBoost), and neural networks. The experimental evaluation confirmed the high accuracy of these forecasting methods. The method for predicting the number of days with atmospheric visibility in Warsaw, Poland, based on meteorological and air quality data is proposed in Majewski et al. (2021). The approach uses the Iman-Conover method to simulate a time series of meteorological and air quality parameters. Statistical verification showed that the model effectively predicts the number of days in a month with visibility within specific ranges. In Kujawska et al. (2022), various machine learning methods for predicting PM_{10} concentrations in Lublin, Poland, are discussed. Methods evaluated include linear regression (LR), k-nearest neighbors regression (k-NNR), support vector machine (SVM), regression trees (RT), Gaussian process regression (GPR), artificial neural networks (ANN), and long short-term memory networks (LSTM). The experimental analysis, based on the coefficient of determination (R^2), indicated that the ANN achieved the best prediction results among the methods tested. Additional studies on air quality analysis in Poland can be found in Gaj et al. (2020) and Kryza et al. (2020).

In He et al. (2022), the forecasting of $PM_{2.5}$ concentration using ANN is considered. Analysis of data from Liaocheng from 2014 to 2021 showed a gradual decrease in $PM_{2.5}$, and the air quality in 2020 was clearly better than in 2019. The ANN model consisted of an input layer (11 parameters), a hidden layer (6 neurons), and an output layer. Training was carried out on 80% of the data, validation on 10%, obtaining high correlation coefficients. The best forecasting results were achieved with the Bayesian regularization algorithm. The results confirm that ANN effectively identifies nonlinear relationships and is an effective tool for predicting monthly $PM_{2.5}$ concentrations. The application of ANN to forecast the hourly concentration of $PM_{2.5}$ and PM_{10} in Chongqing was also described in Guo et al. (2023b).

The work (Guo et al. 2023a) examines the use of ANN and wavelet transform neural networks (WANN) to forecast daily $PM_{2.5}$ concentration in Shanghai. $PM_{2.5}$ concentration decreased by 39.3% from 2014 to 2020. The model is based on the correlation of $PM_{2.5}$ with meteorological data and concentrations from the previous 3 days, identifying the minimum and maximum temperature and maximum atmospheric pressure as key factors. The WANN with the tansig-purelin activation function achieved the best results during all stages of analysis, confirming its effectiveness in identifying the relationship between input and output data. The results suggest that WANN is an accurate tool for $PM_{2.5}$ forecasting and can support air pollution management strategies.

In Guo et al. (2023c), the application of an adaptive model combining ANN with wavelet analysis (WANN) for forecasting daily $PM_{2.5}$ concentrations based on remote sensing data and ground observations is presented. The evaluation was performed using the Pearson correlation coefficient (R), MAPE, RMSE, and MAE. The WANN model performed significantly better than the ANN. The results show that the WANN has higher forecast accuracy, better stability, and smaller error during the training and validation periods, which confirms the effectiveness of the WANN in forecasting the $PM_{2.5}$ concentrations 1 day ahead.

In Guo et al. (2020), the analysis of the forecasting of the air pollution index (API) in Xi'an and Lanzhou using correlation analysis and wavelet transform neural networks (ANN) is presented. The optimal model included the API of the previous 3 days and 16 selected meteorological factors. The API was found to be most strongly correlated with the previous day's value, as well as with the mean, minimum, maximum temperature, and water vapor pressure. The ANN and WANN models trained with the Bayesian regularization algorithm effectively reproduced the API, with WANN achieving better results ($R = 0.8846$ for Xi'an and $R = 0.8906$ for Lanzhou) than ANN. The results show that WANN is effective in short-term API forecasting due to its ability to identify nonlinear relationships and historical patterns. The study can support the development of environmental management policies.

Proposed methods

In this section, we describe the architectures of the four neural networks that one may use for forecasting. Discussed networks include xLSTM, KAN, TCN, and VAE. Firstly, let us introduce the background of the neural networks' work.

Neural networks—the background

Forecasting with neural networks utilizes their ability to uncover intricate patterns in data, enabling accurate predictions of future values based on historical trends. These models learn by analyzing large datasets, iteratively adjusting their parameters to reduce forecast errors and improve precision. The training process begins by feeding input data into the model, which is processed through layers such as convolutional that extract key features. These features are then refined and interpreted through activation functions and fully connected (linear) layers. The network's output is compared to the actual values, and the resulting error is minimized by fine-tuning the model's weights and parameters, ultimately enhancing its predictive accuracy.

The key concepts related to neural network operations can be summarized as follows: The convolutional layer plays a crucial role in detecting local patterns in data, making it indispensable for tasks like image analysis and time series processing. This layer uses a filter (or weight matrix) that moves across the input data, identifying features. Parameters like dilation rate extend the filter's range, capturing broader patterns without adding extra parameters, while the kernel size determines the filter's dimensions, influencing the resolution of detected details. Padding involves adding values (typically zeros) to the input, ensuring the convolutional filter can analyze the entire sequence, including its edges, and preserving the input's original length. Fully connected layers link every neuron to all neurons in the previous layer, enabling the integration of information across processing stages. Activation functions, such as the rectified linear unit (ReLU), allow the network to model non-linear relationships by transmitting signals based on their positivity. The scaled exponential linear unit (SELU) stabilizes training by normalizing activation values. Dropout, a regularization method, deactivates random neurons during training to reduce overfitting. Lastly, the learning rate governs the magnitude of weight updates during optimization, ensuring efficient and stable training.

Extended long short-term memory

LSTM (long short-term memory) is a type of recurrent neural network (RNN) designed specifically to handle long data sequences and memorize information for longer periods of time (Beck et al. 2024). Unlike traditional RNNs, LSTM can better handle the so-called "vanishing gradient" problem, meaning that it can effectively learn dependencies over longer stretches of sequence.

LSTM consists of units called memory cells, which contain the internal state of the cell ("cell state") and three different gates, which include the input gate, forget gate, and output gate. The input gate determines which new information should be added to the cell state. It contains a sigmoid function that decides which values will be updated and a hyperbolic tangent (tanh) function that creates a vector of new candidate values to add to the cell state. The forget gate decides which information from the previous state of the cell will be forgotten and uses a sigmoid function to determine how much each piece of information should be retained or forgotten. The output gate decides what part of the cell state should be the output of the given LSTM unit. This gate, like the previous ones, uses a sigmoid function to decide which information will pass on and a tanh function to set the output.

An extension of the standard LSTM is the xLSTM network, which introduces additional mechanisms or modifications that aim to improve performance or allow better modeling of specific types of data. It can include various variants and additions, such as additional gates, modified activation functions, or special memory mechanisms. The aim of these extensions is to improve the model's ability to cope with specific problems. We propose that, unlike the standard LSTM network, which contains one layer, we create an xLSTM network structure that has four layers. The processed data passes through LSTM layers, where the output from one layer is passed as input to the next. The final output is passed through the linear layer (fully connected—fc). The network structure is thus as follows:

(1) Four layers, each with 64 LSTM units;
(2) A linear layer (fully connected) that matches the dimensionality of the hidden state to match the output size.

The proposed model consists of four LSTM layers, each processing sequential data, storing information about previous time steps, and enabling prediction based on previous data. The first LSTM layer takes input data of size that can be defined by a parameter, and each subsequent layer takes data processed by the previous layer. Hidden state and cell state store information from previous time steps and help the model to "remember" previous data while processing sequences. Hidden states and cells are updated as data flows through the layer. The last layer is a linear layer (fully connected) that maps the output of the last time step, returning the final prediction result.

The network processes a time series by using a sliding window approach, where the time series data is segmented into overlapping windows of a specified size. Each window captures a segment of the time series, and these segments are sequentially fed into the network's layers. The network then processes the data within each window and produces a prediction based on the values contained in that window. The window size is a parameter that can be adjusted according to the specific requirements of the analysis.

Kolmogorov-Arnold network

The Kolmogorov-Arnold networks (KAN) were first introduced in 2024 (Liu et al. 2024). This type of network is composed of fully connected layers, meaning that each neuron in a given layer is connected to every neuron in the subsequent layer. These fully connected layers apply linear transformations to the input data, with each layer learning a set of weights and biases during the training process. These learned parameters are then used to transform the input vectors.

To capture more complex patterns in the data, each layer's output is passed through a non-linear activation function, typically the sine function in the case of KAN. This non-linear activation enables the network to learn intricate relationships within the data. We propose the following structure for the KAN network:

(1) First fully connected layer (FC1) with 64 neurons with sine activation function and dropout;
(2) Second fully connected layer (FC2) with 128 neurons with sine activation function and dropout;
(3) Third fully connected layer (FC3) with 256 neurons with sine activation function and dropout;
(4) Fourth fully connected layer (FC4) with 512 neurons with sine activation function and dropout;
(5) The last layer takes as an input the output of FC4 and produces the final prediction.

Such architecture allows the capturing of more complex patterns in the input data compared to simpler networks.

The network takes a time series as input, processing it with a sliding window of a defined size. The values from subsequent windows are passed through network layers that gradually learn more complex patterns in the data. Finally, the network generates a prediction based on the values from the window. The size of the window can be defined arbitrarily.

Temporal convolutional network

Temporal convolutional networks (TCNs) are specialized for processing time-varying data inputs. A common approach to handling such data involves the use of an entry window and a prediction window, a method that has also been utilized in video-based action segmentation (Lea et al. 2017, 2016). TCNs typically function in two main stages. First, they compute low-level features through a convolutional neural network (CNN), which encodes both spatial and temporal information. These low-level features are then passed into a classifier that captures higher-level temporal patterns, often using techniques like RNNs. TCNs are widely used in various forecasting applications due to their ability to model complex temporal dependencies effectively.

The structure of TCNs is kind of similar to CNNs. We propose to use a TCN with six convolutional layers, which may be summarized as follows:

(1) First 1D-convolution layer with kernel size equal 3, SELU as an activation function and dropout;
(2) Second 1D-convolution layer with dilation rate equal 2, kernel size equal 3, SELU as an activation function and dropout;

(3) Third 1D-convolution layer with dilation rate equal 4, kernel size equal 3, SELU as an activation function and dropout;
(4) Fourth 1D-convolution layer with dilation rate equal 8, kernel size equal 3, SELU as an activation function and dropout;
(5) Fifth 1D-convolution layer with dilation rate equal 16, kernel size equal 3, SELU as an activation function and dropout;
(6) The last convolutional layer reduces the number of channels to the desired output using a 1×1 convolutional operation. The output of the networks is a tensor that contains predicted values for each sequence.

We propose to use a non-linear SELU function (Klambauer et al. 2017), which is a variant of the ReLU function (Nair and Hinton 2010) and is defined as presented as Eq. 1:

$$\mathrm{SELU}(x) = \lambda \begin{cases} x & \text{for } x > 0 \\ \alpha \exp(x) - \alpha & \text{for } x \leq 0 \end{cases} \tag{1}$$

where $\alpha \approx 1.6733$ and $\lambda \approx 1.0507$. Dropout regularization is employed to prevent overfitting by randomly setting a fraction of the activations to zero during training. Additionally, the dilation rate in a convolutional layer specifies how many input values are skipped between each step of the convolution. A dilation rate of 1 corresponds to a standard convolution, whereas higher rates result in skipping more values between the convolutions. Similar to previous networks, the proposed TCN processes the input data with a sliding window of a defined size in order to generate a prediction.

Variational autoencoder

VAE is a generative probabilistic model that is used to learn a representation of data and generate new data that is similar to the training data (Kingma and Welling 2013). VAEs are usually used for different tasks, such as data generation, dimensionality reduction, anomaly detection, interpolation, data modeling, or forecasting (Cai et al. 2023). A VAE consists of three main parts: an encoder, which transforms the input data into a latent space (a space of lower dimensionality); a reparameterization layer that allows us to obtain a sample from a normal distribution, which is defined by the mean and the logarithm of the variance (logvar); and a decoder, which transforms the latent space back to the input space. The latent space is the space where we want to reduce our input data to a lower number of dimensions while retaining the most important information. In a VAE, instead of uniquely encoding each input data into a single point, we learn a probability distribution that represents where the data might lie in the latent space. The proposed VAE can be characterized as follows:

(1) Encoding part:
- Linear layer fc1 (dense) transforming input data to the dimension of the hidden vector **h**;
- Linear layer fc2 (dense) which
  (a) transforms the vector **h** by determining the vector of means **m** for each dimension of the latent space;
  (b) transforms the vector **h** by determining the vector of logarithms of variances (logvar) **l** for each dimension of the latent space;

(2) Reparametrization:
- Generates the vector **z** which is the representation in the latent space;

(3) Decoding part:
- Linear layer fc3 (dense) transforming the latent space to the dimension of the hidden vector **h**;
- Linear layer fc4 (dense) transforming vectors from hidden layer to input dimension.

The operation of the encoder can be described as follows: The input data (e.g., a vector of values) is fed to the fc1 layer, which transforms it into a hidden vector **h**. This vector is a kind of abstract representation of the input data after initial transformation by the encoder. The vector **h** is then transformed into two vectors: the mean **m** and the logarithm of the variance **l**, which define the normal distribution. The vector **m** represents the mean value of the normal distribution to which the input data will be mapped. The vector **l** (logarithm of the variance) is a vector that represents the logarithm of the variance of this distribution. The variance defines the spread of the data around the mean, i.e., how much the samples in the latent space can deviate from the mean. So the vectors **m** and **l** define the normal distribution, which contains the encoded representation of the input data. This allows for probabilistic modeling of the data, which allows the VAE to generate new samples that may look similar to the input data.

Reparameterization enables the gradient to propagate through the stochastic layer. After obtaining vectors **m** and **l**, we want to obtain a sample from the normal distribution that these parameters define. With reparameterization, we can use optimization techniques like backpropagation to train the model. The vector **z** is generated from the normal distribution $\mathcal{N}(0, 1)$ by using vectors **m** and **l**. More precisely, the **z** is calculated in the following manner:

$$\mathbf{z} = \mathbf{m} + \mathrm{std} \times \epsilon \tag{2}$$

where std $= \exp(0.5 \times \mathbf{l})$, and ϵ is a sample from the normal distribution.

The decoder operation is to transform the sample from the latent space using the decoder (fc3 and fc4) into a reconstruction of the input data.

The VAE can be used to generate data (prediction) in the following way. After training is complete, one can generate new data (vector) $\mathbf{z}$ by sampling from the standard normal distribution $\mathcal{N}(0, 1)$ from the latent space. Therefore, each element of the vector $\mathbf{z}$ is independently drawn from a normal distribution with mean 0 and standard deviation 1. This vector is then fed to the decoder, which is tasked with transforming this vector into data in the original space. Thus, this sample, when decoded into the input space, yields new, generated data. In the context of prediction, the generated data can be treated as predictions. This works because the model has learned to map the input data to the latent space such that these distributions are close to $\mathcal{N}(0, 1)$, so we get samples in the latent space that should correspond to reasonable data in the input space after being transformed by the decoder.

Experimental evaluation

In this section, we benchmark the proposed methods by conducting forecasting simulations using real data from the Chief Inspectorate of Environmental Protection (GIOS). The data from GIOS consist of concentration values (measured in $\mu g/m^3$) for various air pollutants. Specifically, our objective is to train the proposed methods (xLSTM, KAN, TCN, and VAE) using data that represent different concentrations of air pollutants from several Polish cities (referred to as the original data). We then compare the results generated by the proposed methods with the original data to assess the forecasting accuracy and overall quality of these methods. This comparison enables a thorough evaluation of the effectiveness of the forecasting models we propose.

Collected data and area description

The forecasting evaluation was conducted with the PM_{10} (particulate matter) and its following components: arsenic (As), cadmium (Cd), nickel (Ni), and lead (Pb). These pollutants were analyzed in the following air quality monitoring stations in Poland: DsWalbrzWyso (Wałbrzych), KpToruDziewu (Toruń), LuZaganKochaMOB (Żagań), MpKrakBulwar (Kraków), PkMielPogodn (Mielec), SlKatoKossut (Katowice), WmPuszczaBor (Puszcza Borecka), and ZpSzczAndrze (Szczecin). These station codes, along with the city names in parentheses, are used throughout this paper. The selection of the listed cities and the air pollutants was based on the availability and consistency of high-quality historical data from monitoring stations. These cities represent a mix of urban, industrial, and suburban areas, capturing a range of pollution levels typically observed across Poland. For instance,

- Urban areas like Kraków experience high pollution levels due to traffic and dense population;
- Industrial regions such as Katowice are characterized by elevated emissions from industrial activities;
- Smaller towns like Żagań, Mielec, or cities in the northern part of Poland like Szczecin or Toruń have lower pollution levels, providing contrast to the urban and industrial settings.

Additionally, the listed air pollutants have significantly different concentration levels, allowing for a more comprehensive evaluation of the proposed methods. The locations of the air quality monitoring stations are shown in Fig. 1.

We present the forecasting results based on daily data collected from January 1, 2022, to December 31, 2022, for all listed air pollutants. Consequently, the input data for each of the proposed methods consists of 365 daily observations. The data used for training the proposed methods has not been pre-processed or normalized; instead, the models were trained directly on the original data. This decision was made to ensure the proposed forecasting methods could operate on real-world data without relying on pre-treatment steps that might not be feasible in practical applications. By using unprocessed data, we aimed to assess the robustness and adaptability of our models in handling inherent variations and noise present in the dataset.

Usage of the proposed methods, hyperparameters, and implementation

The proposed methods were employed in the following manner: each model (except VAE) was configured to accept multiple input values and produce a single output value, which corresponds to the prediction. The number of input values, referred to as the "window," varied across different experiments. Specifically, each network (xLSTM, KAN, and TCN) generated predictions using windows of 3, 5, and 7 input values. The window of a defined size moved over the input data, and the network produced one prediction; after this, the window was moving over the next position of the input data, to make the network produce the next prediction, and so on to the end of the input data. The only exception was the VAE, which was utilized in just one configuration. As a result, using data from 5 air pollutants across 8 air quality monitoring stations, forthe 4 proposed methods generated a total of 400 series of predictions.

To evaluate the performance of the proposed methods, we compared their predictions not only with the original data but also with forecasts generated by several well-known state-

Fig. 1 The location of the air quality monitoring stations in Poland used in the experiment

of-the-art algorithms: exponential smoothing, Holt-Winters exponential smoothing, autoregressive integrated moving average (ARIMA), and random forest. Exponential smoothing is a simple yet effective technique for forecasting time series data by weighting recent observations more heavily, making it suitable for short-term predictions. Holt-Winters exponential smoothing extends this approach by utilizing seasonality and trends, making it particularly relevant for capturing cyclical patterns in air quality data. ARIMA is a widely applied statistical method for time series analysis, offering a benchmark for assessing the improvements provided by deep learning models in handling temporal dependencies. Random forest, on the other hand, represents a traditional machine learning approach that has demonstrated effectiveness in various predictive tasks, including air pollution forecasting, due to its ability to handle nonlinear relationships and high-dimensional data. These algorithms were chosen because they are well-established in the field, allowing us to demonstrate the advantages of our deep learning-based methods in terms of accuracy and robustness. As these algorithms are well-known, they are not described in detail within this paper.

The hyperparameters for all the proposed networks were determined experimentally and are generally consistent across the models. Most networks were trained for 1000 epochs using the Adam optimizer, with a learning rate set to 0.001. The exception was the VAE, which was trained for 100 epochs. Dropout was applied to prevent overfitting: KAN used a dropout rate of 0.3, while TCN used 0.2. The xLSTM and VAE models did not utilize dropout. The size of the latent vector $\mathbf{z}$ in VAE was equal to the size of the input data, i.e., 365. The key hyperparameters are summarized in Table 1.

Table 1 Hyperparameters of the proposed networks

	xLSTM	KAN	TCN	VAE
Input (# of values)	3/5/7	3/5/7	3/5/7	—
Output (# of values)	1	1	1	365
Dropout	not used	0.3	0.2	not used
Learning rate	0.001	0.001	0.001	0.001
Optimizer	Adam	Adam	Adam	Adam
# of epochs	1000	1000	1000	100

Fig. 2 Arsenic forecasts compared with the original data (SlKatoKossut)

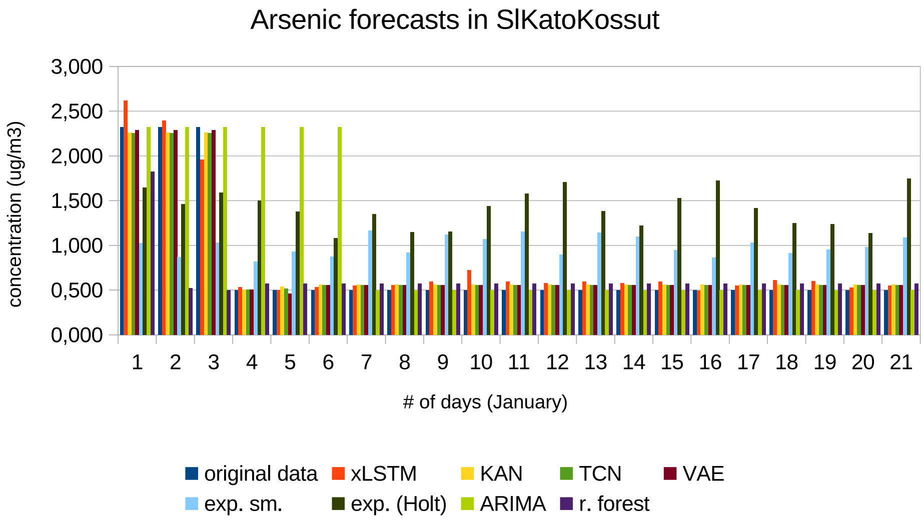


Fig. 3 Cadmium forecasts compared with the original data (SlKatoKossut)

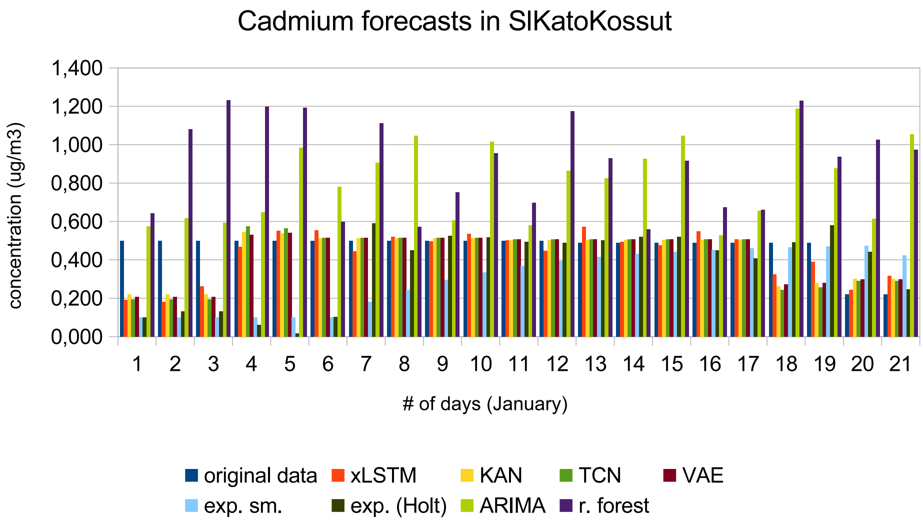
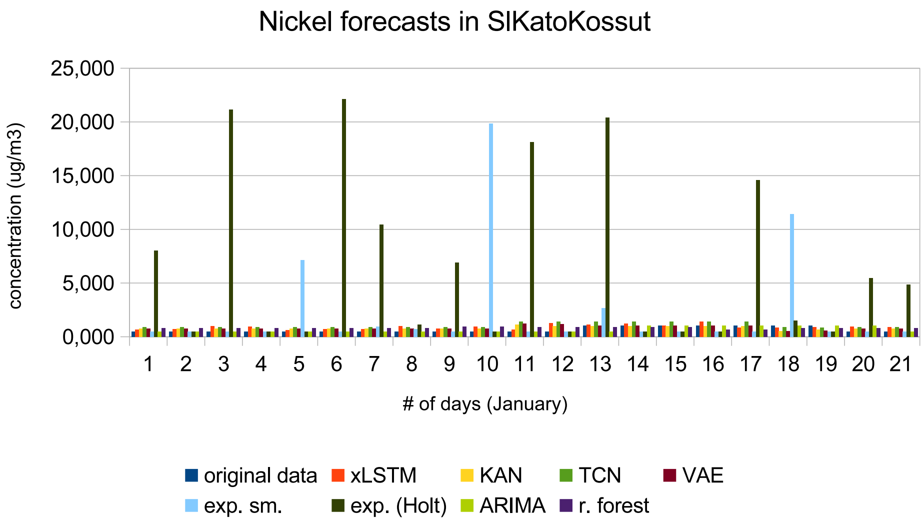


Fig. 4 Nickel forecasts compared with the original data (SlKatoKossut)



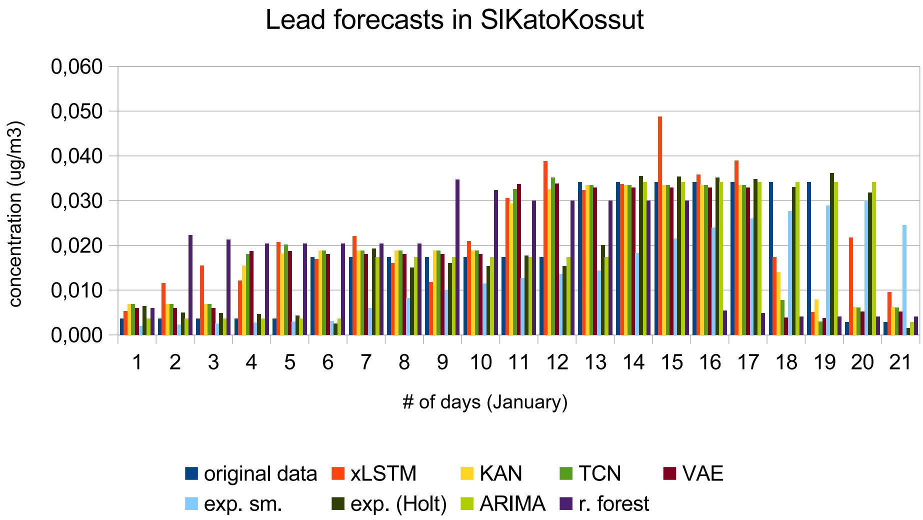


Fig. 5 Lead forecasts compared with the original data (SlKatoKossut)

The scripts for the proposed methods and state-of-the-art algorithms were implemented under Python programming language with PyTorch framework with Nvidia CUDA support. The scripts for statistical verification were implemented in Matlab. Experiments were held on a personal computer equipped with an Intel Core i5-10400@2.9GHz CPU with 32 gigabytes of RAM and a Nvidia RTX 3060 GPU with 12 gigabytes of video memory.

Forecasting examples—visual inspection

Due to the very large number of forecast models (400 figures), we skip presenting them in the paper. We only show partial forecasts for all air pollutants in SlKatoKossut that may be seen in Figs. 2, 3, 4, 5, and 6.

Visual inspection confirms the reliability of the proposed methods. For instance, the forecasts of lead concentrations at the SlKatoKossut air monitoring station generated by the xLSTM model almost perfectly overlap with the original data, demonstrating a high level of accuracy. A similar observation is made for cadmium predictions using the VAE model, where the forecasted values closely match the actual recorded concentrations. These visual confirmations support the effectiveness of the proposed forecasting models. On the other hand, very low prediction accuracy is observed in the case of Holt exponential smoothing for the concentrations of the nickel. Detailed numerical analysis of the forecast models, including error metrics and statistical evaluations, is presented in the subsequent subsections.

Statistical verification

Statistical verification through hypothesis testing is crucial in determining whether the analyzed data sets differ significantly from each other. In this study, we specifically assessed whether the original data from GIOS significantly differed

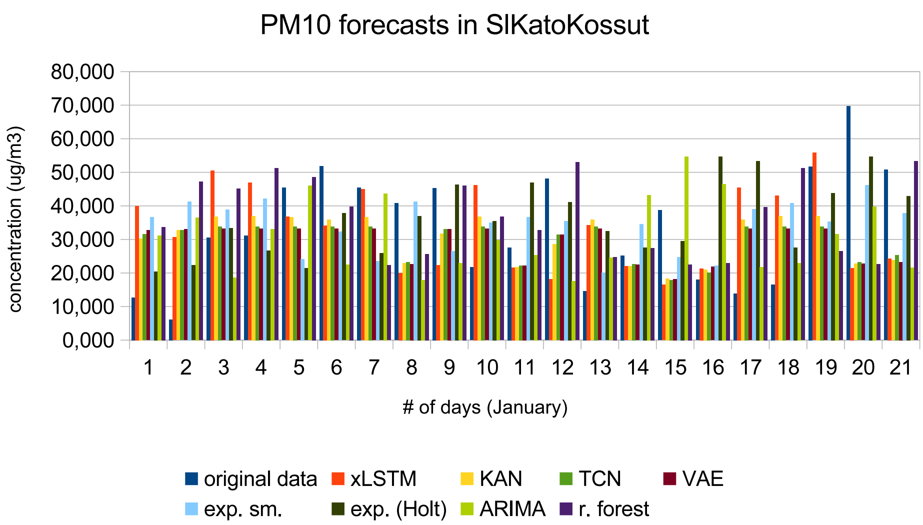


Fig. 6 PM_{10} forecasts compared with the original data (SlKatoKossut)

from the forecasts generated by the proposed deep learning methods (xLSTM, KAN, TCN, and VAE). Additionally, we compared our predictions with the forecasts produced by established state-of-the-art algorithms such as exponential smoothing, Holt-Winters exponential smoothing, ARIMA, and random forest. All hypothesis tests were conducted at a significance level of $\alpha = 0.05$. The results of these tests provide insights into the accuracy and reliability of the proposed methods in comparison to existing forecasting techniques.

Normality tests

To assess the normality of the analyzed data, we employed the Shapiro-Wilk test, a widely used method for checking whether a sample comes from a normally distributed population (Shapiro and Wilk 1965). The hypotheses are defined as follows:

- $\mathcal{H}_0$: The analyzed data follows the standard distribution;
- $\mathcal{H}_1$: The analyzed data does not follow the standard distribution.

To accept or reject the null hypothesis, one should calculate the Shapiro-Wilk test statistic S and then calculate the p-value (let us further name it shortly p). The p is compared with the significance level, and if greater, then there is no reason to reject the null hypothesis $\mathcal{H}_0$. Otherwise, the alternative hypothesis $\mathcal{H}_1$ is accepted. Alternatively, if S is greater than the test critical value, the $\mathcal{H}_0$ might be accepted.

We examined whether the data followed a normal distribution, using a critical value of 0.07 for the 365 input samples. The shortened results are presented as Table 2. The complete results, including p-values and test statistics S, are detailed in the Appendix (Tables 10, 11, 12, 13, 14, 15, 16, and 17).

The analysis reveals that the null hypothesis $\mathcal{H}_0$ should be rejected in nearly all cases, as the p-values are consistently around 0.00 and the test statistics S are below the critical value. Therefore, the data do not adhere to a normal distribution, necessitating the use of the non-parametric Kruskal-Wallis ANOVA test for subsequent analysis.

Kruskal-Wallis ANOVA

The non-parametric Kruskal-Wallis ANOVA, or Kruskal-Wallis one-way analysis of variance for ranks, is used to compare more than two independent samples to determine if there are significant differences between their medians. This test is suitable for data that do not meet the assumptions of normality required by parametric tests. It extends the U Mann–Whitney test to more than two groups. This test is used for the verification of the insignificance of differences of the medians in ($k > 2$) populations (Kruskal and Wallis 1952; Kruskal 1952). The hypotheses for k groups are defined as follows:

- $\mathcal{H}_0$: $\theta_1 = \theta_2 = \cdots = \theta_k$ (all the medians are equal);
- $\mathcal{H}_1$: Not all medians θ_j are equal ($j = 1, 2, \ldots, k$).

To determine whether to accept or reject the null hypothesis, the test statistic H must first be calculated, followed by the computation of the p-value. This p-value is then compared to the significance level; if it is greater than the significance level, the null hypothesis $\mathcal{H}_0$ cannot be rejected. Conversely, if the p-value is less, the alternative hypothesis $\mathcal{H}_1$ is supported. Alternatively, one can compare the test statistic H directly with the critical value. If H exceeds the critical value, the null hypothesis $\mathcal{H}_0$ should be rejected.

Experiment I—forecasts comparison with original data

In this analysis, we perform a statistical comparison of the medians between the original data and the forecasts produced by the proposed methods. With a total of 11 groups under consideration, the degrees of freedom between these groups are calculated as $11 - 1 = 10$. The degrees of freedom within the groups amount to 4459. Consequently, the critical value for the Kruskal-Wallis ANOVA test is set at 1.83. To validate the null hypothesis $\mathcal{H}_0$, the test statistic H must not exceed this critical value. The detailed results of the Kruskal-Wallis ANOVA test are provided in Table 3.

Results indicate that in the majority of cases, the values of the test statistic do not exceed the critical value of the test, which allows for accepting the hypothesis $\mathcal{H}_0$. However, in some cases, the test statistic H is significantly greater than the critical value. Such a situation is observed for cadmium (Cd) (ZpSzczAndrze) and PM_{10} (DsWalbrzWyso, KpToruDziewu, MpKrakBulwar, PkMielPogodn, and SlKatoKossut). Therefore, it means that in these cases, we cannot accept the $\mathcal{H}_0$, so the data differ significantly from each other. Checking the detailed tendencies may be realized with the post hoc analysis. However, we can assume that in most cases, the forecasts are accurate for almost all air quality monitoring stations.

Experiment II—forecasts comparison with original data and state-of-the-art algorithms

In this study, we conducted a statistical comparison between the medians of the original data and the forecasts produced by both the proposed methods and state-of-the-art algorithms. We examined 15 distinct groups in this analysis, where the degrees of freedom between groups amount to $15 - 1 = 14$, and the degrees of freedom within groups total 5831. Consequently, the critical value for the Kruskal-Wallis ANOVA test is 1.67. The results of this test are detailed in Table 4.

Table 2 Hypotheses of the normality tests

	original	xLSTM 3	xLSTM 5	xLSTM 7	KAN 3	KAN 5	KAN 7	TCN 3	TCN 5	TCN 7	VAE	exp. sm.	exp. (Holt)	ARIMA	r. forest
DsWalbrzWyso															
As(PM_{10})	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cd(PM_{10})	1	0	0	0	0	0	0	1	0	1	1	0	0	0	1
Ni(PM_{10})	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
Pb(PM_{10})	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1
PM_{10}	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1
KpToruDziewu															
As(PM_{10})	1	1	1	1	1	1	0	1	0	1	0	0	0	0	1
Cd(PM_{10})	1	0	1	0	1	1	1	0	1	0	0	0	1	1	1
Ni(PM_{10})	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Pb(PM_{10})	0	0	1	1	0	0	0	0	0	0	0	1	0	0	1
PM_{10}	0	1	1	0	0	0	0	0	1	0	0	1	0	0	1
LuZaganKochaMOB															
As(PM_{10})	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
Cd(PM_{10})	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ni(PM_{10})	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1
Pb(PM_{10})	0	0	0	0	1	1	1	1	1	1	1	0	1	0	0
PM_{10}	0	1	1	1	1	1	1	1	1	1	1	0	1	0	1
MpKrakBulwar															
As(PM_{10})	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1
Cd(PM_{10})	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ni(PM_{10})	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Pb(PM_{10})	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PM_{10}	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1
PkMielPogodn															
As(PM_{10})	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1
Cd(PM_{10})	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ni(PM_{10})	0	1	1	0	1	1	1	1	1	1	1	0	0	0	1
Pb(PM_{10})	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PM_{10}	1	0	0	1	1	1	1	1	1	1	1	0	0	0	1

Table 2 continued

	original	xLSTM 3	xLSTM 5	xLSTM 7	KAN 3	KAN 5	KAN 7	TCN 3	TCN 5	TCN 7	VAE	exp. sm.	exp. (Holt)	ARIMA	r. forest
SlKatoKossut															
$As(PM_{10})$	1	0	1	1	0	1	0	1	1	1	1	0	0	1	0
$Cd(PM_{10})$	0	0	0	0	1	1	1	0	0	0	0	1	0	1	0
$Ni(PM_{10})$	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1
$Pb(PM_{10})$	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1
PM_{10}	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
WmPuszczaBor															
$As(PM_{10})$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$Cd(PM_{10})$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$Ni(PM_{10})$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$Pb(PM_{10})$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PM_{10}	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ZpSzczAndrze															
$As(PM_{10})$	0	1	0	0	0	1	0	0	0	0	0	1	1	1	1
$Cd(PM_{10})$	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1
$Ni(PM_{10})$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$Pb(PM_{10})$	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1
PM_{10}	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

0 denotes that data represent normal distribution (accepting $\mathcal{H}_0$), and 1 denotes the difference from the normal distribution (accepting $\mathcal{H}_1$). The number next to the name of the proposed method denotes the size of the window (e.g., TCN 3 denotes the TCN with a window of size 3; KAN 7 denotes the KAN with a window of size 7); exp. sm. denotes exponential smoothing algorithm; exp. (Holt) denotes Holt-Winters exponential smoothing algorithm; ARIMA stands for autoregressive integrated moving average algorithm; r. forest denotes the random forest algorithm

Table 3 Values of test statistic H of the Kruskal-Wallis ANOVA test—forecasts comparison with original data

	As(PM$_{10}$)	Cd(PM$_{10}$)	Ni(PM$_{10}$)	Pb(PM$_{10}$)	PM$_{10}$
DsWalbrzWyso	0.26	0.11	0.15	0.16	2.09
KpToruDziewu	0.08	0.64	0.20	0.20	1.57
LuZaganKochaMOB	0.11	0.28	0.14	0.73	0.18
MpKrakBulwar	0.18	0.23	0.59	1.02	6.51
PkMielPogodn	0.25	0.47	0.33	0.78	6.39
SlKatoKossut	0.09	0.14	0.19	0.13	3.38
WmPuszczaBor	0.16	0.34	0.46	1.03	0.41
ZpSzczAndrze	0.34	2.72	0.14	0.44	0.44

The obtained results suggest that in the majority of the cases, the test statistic H is greater than the critical value (1.49). The hypothesis $\mathcal{H}_0$ might be accepted only on cases nickel (Ni) and lead (Pb) (KpToruDziewu); cadmium (Cd), nickel (Ni), and lead (Pb) (LuZaganKochaMOB); lead (Pb) (MpKrakBulwar); nickel (Ni) (PkMielPogodn); cadmium (Cd) and lead (Pb) (SlKatoKossut); arsenic (As), cadmium (Cd), nickel (Ni), and lead (Pb) (WmPuszczaBor); and lead (Pb) (ZpSzczAndrze). The most challenging was PM$_{10}$, where in all cases the alternative hypothesis $\mathcal{H}_1$ should be accepted. To check the specific tendencies, we can perform the post hoc analysis.

Due to the fact that Kruskal-Wallis ANOVA only generally indicates, whether the groups differ (or not) significantly without pointing precise differences, the post hoc analysis may be calculated in cases, where the $\mathcal{H}_0$ is rejected. The post hoc indicates the sum of the ranks of the particular groups, so one may deduce the precise tendencies between them. In Table 5, the values of the sum of ranks for the particular data are described.

Results indicate that the sum of ranks between the original input data and the forecasts generated by the proposed methods are similar in the case of SlKatoKossut, MpKrakBulwar, and WmPuszczaBor. However, the sum of ranks for nearly all air quality stations except WmPuszczaBor for state-of-the-art algorithms is greater than the original data, which suggests that their predictions are less accurate. The sum of ranks for the proposed methods is in general similar to the original data, which confirms their accuracy. To conclude, the overall number of forecast models generated by the proposed methods was equal to 400; the number of forecast models by state-of-the-art algorithms was 160. The number of forecasts generated by the proposed methods that were statistically different from the original data was 6 (1%); 67 forecast models (58%) were significantly different by the algorithms from the literature. This clearly confirms the high accuracy of the proposed methods.

Results of error measures

To evaluate the performance of the forecasts generated by both the proposed and state-of-the-art methods, we calculate several key error metrics: MAE (Eq. 3) and RMSE (Eq. 4).

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - x_t}{n}\right| \tag{3}$$

where x_t is the actual value, y_t is a predicted value, and n is the number of observations.

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} \tag{4}$$

where x_t is the actual value, y_t is a predicted value, and n is the number of observations, and

$$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n}(x_t - y_t)^2 \tag{5}$$

Table 4 Values of test statistic of the Kruskal-Wallis ANOVA test—forecasts comparison with original data and state-of-the-art algorithms

	As(PM$_{10}$)	Cd(PM$_{10}$)	Ni(PM$_{10}$)	Pb(PM$_{10}$)	PM$_{10}$
DsWalbrzWyso	483.47	1192.96	122.76	963.41	15.65
KpToruDziewu	51.87	42.91	1.43	0.18	13.47
LuZaganKochaMOB	2.28	0.44	0.96	0.53	3.45
MpKrakBulwar	33.69	39.21	4.76	0.74	41.09
PkMielPogodn	2.85	5.51	0.45	0.78	13.89
SlKatoKossut	5.82	1.13	6.10	0.15	62.01
WmPuszczaBor	0.23	0.27	0.47	1.04	8.61
ZpSzczAndrze	17.94	2.50	3.65	0.34	8.34

Table 5 Post hoc analysis of the sum of ranks for Experiment I and II

	original	proposed	state-of-the-art
DsWalbrzWyso	2084	2418	4652
KpToruDziewu	2338	2764	3551
LuZaganKochaMOB	2354	2930	3047
MpKrakBulwar	2550	2785	3434
PkMielPogodn	2156	2923	3118
SlKatoKossut	2669	2858	3185
WmPuszczaBor	2727	2947	2906
ZpSzczAndrze	2078	3002	2902

Due to the large number of evaluation tables, we provide them in the Appendix for clarity. The tables detail the MAE for each air pollutant at every monitoring station, as shown in Tables 18, 19, 20, 21, 22, 23, 24, and 25. Similarly, Tables 26, 27, 28, 29, 30, 31, 32, and 33 present the RMSE values. The shortened tables with mean values for MAE and RMSE are presented as Tables 6 and 8.

MAE analysis

In Table 6, the mean values of MAE may be observed, both for all the proposed methods and state-of-the-art algorithms. In Table 7, the MAE values per each air pollutant are presented.

For all cases, the proposed methods achieved lower MAE values than state-of-the-art algorithms. According to the minimum and maximum values of the concentration of the particular air pollutants, we may find such results satisfactory. For the proposed methods, the MAE values obtained 2.58, maximally (for SlKatoKossut), while algorithms from the literature mostly exceeded the value of 2.00, usually reaching up to 3.50–4.00. The highest MAE value for state-of-the-art is observed for SlKatoKossut (4.13). Lower MAE value greater accuracy of the proposed methods over the algorithms from the literature. Detailed MAE values for each air pollu-

Table 6 Average MAE error metrics for all air pollutants for the proposed and state-of-the-art methods from all the air quality monitoring stations

	Proposed	State-of-the-art
DsWalbrzWyso	2.41	3.84
KpToruDziewu	2.17	2.64
LuZaganKochaMOB	1.70	2.07
MpKrakBulwar	2.54	3.33
PkMielPogodn	1.49	1.61
SlKatoKossut	2.58	4.13
WmPuszczaBor	1.31	1.47
ZpSzczAndrze	1.54	2.11

Table 7 Average MAE error metrics per each air pollutant; exp. sm. denotes exponential smoothing algorithm; exp. (Holt) denotes Holt-Winters exponential smoothing algorithm; ARIMA stands for autoregressive integrated moving average algorithm; r. forest denotes the random forest algorithm

Method	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}
DsWalbrzWyso					
xLSTM	0.31	0.07	0.24	0.01	11.54
KAN	0.24	0.05	0.24	0.00	9.74
TCN	0.25	0.05	0.24	0.00	9.82
VAE	0.44	0.05	0.24	0.00	10.20
exp. sm	2.69	1.00	1.19	0.15	15.21
exp. (Holt)	1.75	1.01	1.11	0.15	15.00
ARIMA	1.74	1.01	1.37	0.15	14.81
r. forest	1.75	1.01	1.06	0.15	14.46
KpToruDziewu					
xLSTM	0.12	0.03	0.51	0.00	9.74
KAN	0.09	0.02	0.45	0.00	8.52
TCN	0.09	0.02	0.45	0.00	8.61
VAE	0.10	0.02	0.46	0.00	9.20
exp. sm	0.65	0.14	1.35	0.00	10.78
exp. (Holt)	0.60	0.01	0.28	0.00	11.50
ARIMA	0.54	0.10	1.42	0.00	11.86
r. forest	0.48	0.12	1.38	0.00	11.66
LuZaganKochaMOB					
xLSTM	0.73	0.04	0.33	0.00	6.95
KAN	0.65	0.04	0.34	0.00	5.94
TCN	0.66	0.04	0.33	0.00	5.97
VAE	0.65	0.04	0.33	0.00	6.14
exp. sm	2.32	0.06	1.04	0.00	7.56
exp. (Holt)	2.23	0.02	1.07	0.00	8.20
ARIMA	0.00	0.00	1.03	0.00	7.76
r. forest	1.15	0.07	0.51	0.00	8.27
MpKrakBulwar					
xLSTM	0.15	0.08	0.26	0.00	12.16
KAN	0.12	0.06	0.19	0.00	11.25
TCN	0.12	0.06	0.19	0.00	11.16
VAE	0.12	0.06	0.19	0.00	10.92
exp. sm	0.67	0.32	0.49	0.00	14.99
exp. (Holt)	0.65	0.30	0.51	0.00	15.12
ARIMA	0.63	0.25	0.51	0.00	14.67
r. forest	0.64	0.26	0.35	0.01	16.19
PkMielPogodn					
xLSTM	0.07	0.03	0.18	0.00	7.57
KAN	0.05	0.03	0.14	0.00	6.55
TCN	0.05	0.03	0.14	0.00	6.26
VAE	0.05	0.03	0.16	0.00	7.52
exp. sm	0.23	0.15	0.23	0.00	7.92
exp. (Holt)	0.03	0.01	0.09	0.00	9.10

Table 7 continued

Method	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}
ARIMA	0.01	0.01	0.53	0.00	5.65
r. forest	0.07	0.16	0.54	0.00	7.44
SlKatoKossut					
xLSTM	0.23	0.21	1.46	0.01	9.93
KAN	0.19	0.19	1.35	0.01	9.01
TCN	0.19	0.19	1.45	0.01	9.03
VAE	0.19	0.19	1.41	0.01	9.12
exp. sm	0.51	0.28	6.51	0.01	15.07
exp. (Holt)	0.59	0.10	6.70	0.00	15.83
ARIMA	0.08	0.53	0.50	0.00	17.01
r. forest	0.35	0.54	2.26	0.01	15.75
WmPuszczaBor					
xLSTM	0.05	0.02	0.16	0.00	5.71
KAN	0.04	0.02	0.13	0.00	5.35
TCN	0.04	0.02	0.13	0.00	5.38
VAE	0.04	0.01	0.13	0.00	5.62
exp. sm	0.05	0.02	0.20	0.00	6.36
exp. (Holt)	0.02	0.01	0.08	0.00	8.00
ARIMA	0.01	0.01	0.05	0.00	7.25
r. forest	0.07	0.03	0.23	0.00	7.08
ZpSzczAndrze					
xLSTM	0.06	0.00	0.45	0.00	6.69
KAN	0.04	0.00	0.32	0.00	5.65
TCN	0.04	0.02	0.32	0.00	5.74
VAE	0.04	0.00	0.34	0.00	6.46
exp. sm	0.22	0.01	1.32	0.00	9.35
exp. (Holt)	0.22	0.00	1.38	0.00	9.67
ARIMA	0.22	0.00	0.14	0.00	9.09
r. forest	0.20	0.01	1.47	0.00	8.91

tant from every air quality monitoring station may be seen in the Appendix as Tables 18–25.

RMSE analysis

The mean values of RMSE for the proposed methods and algorithms from the literature are shown in Table 8. The RMSE values per each air pollutant are presented in Table 9.

The proposed methods achieved lower values of RMSE measure than state-of-the-art algorithms in all cases. The lowest values are observed for WmPuszczaBor—1.81 for the proposed methods. The highest values for the algorithms from the literature obtained were 4.88, 4.09, and 5.41, for DsWalbrzWyso, MpKrakBulwar, and SlKatoKossut, respectively. Lower RMSE values denote more accurate forecasts.

Table 8 Average RMSE error metrics for all air pollutants for the proposed and state-of-the-art methods from all the air quality monitoring stations

	proposed	state-of-the-art
DsWalbrzWyso	3.70	4.88
KpToruDziewu	3.26	3.49
LuZaganKochaMOB	2.59	2.88
MpKrakBulwar	3.46	4.09
PkMielPogodn	2.12	2.18
SlKatoKossut	3.91	5.41
WmPuszczaBor	1.81	1.95
ZpSzczAndrze	2.21	2.69

Table 9 Average RMSE error metrics per each air pollutant; exp. sm. denotes exponential smoothing algorithm; exp. (Holt) denotes Holt-Winters exponential smoothing algorithm; ARIMA stands for autoregressive integrated moving average algorithm; r. forest denotes the random forest algorithm

Method	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}
DsWalbrzWyso					
xLSTM	0.31	0.07	0.24	0.01	11.54
KAN	0.24	0.05	0.24	0.00	9.74
TCN	0.25	0.05	0.24	0.00	9.82
VAE	0.44	0.05	0.24	0.00	10.20
exp. sm	2.69	1.00	1.19	0.15	15.21
exp. (Holt)	1.75	1.01	1.11	0.15	15.00
ARIMA	1.74	1.01	1.37	0.15	14.81
r. forest	1.75	1.01	1.06	0.15	14.46
KpToruDziewu					
xLSTM	0.12	0.03	0.51	0.00	9.74
KAN	0.09	0.02	0.45	0.00	8.52
TCN	0.09	0.02	0.45	0.00	8.61
VAE	0.10	0.02	0.46	0.00	9.20
exp. sm	0.65	0.14	1.35	0.00	10.78
exp. (Holt)	0.60	0.01	0.28	0.00	11.50
ARIMA	0.54	0.10	1.42	0.00	11.86
r. forest	0.48	0.12	1.38	0.00	11.66
LuZaganKochaMOB					
xLSTM	0.73	0.04	0.33	0.00	6.95
KAN	0.65	0.04	0.34	0.00	5.94
TCN	0.66	0.04	0.33	0.00	5.97
VAE	0.65	0.04	0.33	0.00	6.14
exp. sm	2.32	0.06	1.04	0.00	7.56
exp. (Holt)	2.23	0.02	1.07	0.00	8.20
ARIMA	0.00	0.00	1.03	0.00	7.76
r. forest	1.15	0.07	0.51	0.00	8.27
MpKrakBulwar					
xLSTM	0.15	0.08	0.26	0.00	12.16
KAN	0.12	0.06	0.19	0.00	11.25
TCN	0.12	0.06	0.19	0.00	11.16
VAE	0.12	0.06	0.19	0.00	10.92
exp. sm	0.67	0.32	0.49	0.00	14.99

Table 9 continued

Method	$As(PM_{10})$	$Cd(PM_{10})$	$Ni(PM_{10})$	$Pb(PM_{10})$	PM_{10}
exp. (Holt)	0.65	0.30	0.51	0.00	15.12
ARIMA	0.63	0.25	0.51	0.00	14.67
r. forest	0.64	0.26	0.35	0.01	16.19
PkMielPogodn					
xLSTM	0.07	0.03	0.18	0.00	7.57
KAN	0.05	0.03	0.14	0.00	6.55
TCN	0.05	0.03	0.14	0.00	6.26
VAE	0.05	0.03	0.16	0.00	7.52
exp. sm	0.23	0.15	0.23	0.00	7.92
exp. (Holt)	0.03	0.01	0.09	0.00	9.10
ARIMA	0.01	0.01	0.53	0.00	5.65
r. forest	0.07	0.16	0.54	0.00	7.44
SlKatoKossut					
xLSTM	0.23	0.21	1.46	0.01	9.93
KAN	0.19	0.19	1.35	0.01	9.01
TCN	0.19	0.19	1.45	0.01	9.03
VAE	0.19	0.19	1.41	0.01	9.12
exp. sm	0.51	0.28	6.51	0.01	15.07
exp. (Holt)	0.59	0.10	6.70	0.00	15.83
ARIMA	0.08	0.53	0.50	0.00	17.01
r. forest	0.35	0.54	2.26	0.01	15.75
WmPuszczaBor					
xLSTM	0.05	0.02	0.16	0.00	5.71
KAN	0.04	0.02	0.13	0.00	5.35
TCN	0.04	0.02	0.13	0.00	5.38
VAE	0.04	0.01	0.13	0.00	5.62
exp. sm	0.05	0.02	0.20	0.00	6.36
exp. (Holt)	0.02	0.01	0.08	0.00	8.00
ARIMA	0.01	0.01	0.05	0.00	7.25
r. forest	0.07	0.03	0.23	0.00	7.08
ZpSzczAndrze					
xLSTM	0.06	0.00	0.45	0.00	6.69
KAN	0.04	0.00	0.32	0.00	5.65
TCN	0.04	0.02	0.32	0.00	5.74
VAE	0.04	0.00	0.34	0.00	6.46
exp. sm	0.22	0.01	1.32	0.00	9.35
exp. (Holt)	0.22	0.00	1.38	0.00	9.67
ARIMA	0.22	0.00	0.14	0.00	9.09
r. forest	0.20	0.01	1.47	0.00	8.91

Proposed methods mostly achieved the value RMSE equal to 3.50–4.0. Detailed RMSE values for each air pollutant from every air quality monitoring station may be seen in the Appendix as Tables 26–33.

Discussion

Air pollution forecasting is crucial for protecting public health and the environment by providing early warnings of high pollution periods and enabling preventive measures, such as limiting outdoor activities or imposing temporary emission regulations. Accurate forecasts generated by the proposed models support environmental policies by facilitating proactive actions, such as traffic restrictions, temporary industrial reductions, and the issuance of alerts during pollution episodes. These forecasts also strengthen public health interventions by informing vulnerable populations, including children, the elderly, and individuals with respiratory conditions, about periods of elevated pollution levels, enabling them to take protective measures. On a broader scale, long-term forecasts offer valuable insights for urban planning, such as optimizing green spaces or targeting emission reductions in specific areas. These capabilities underline the societal and environmental importance of forecasting systems, which are essential for improving air quality, protecting public health, and addressing the pressing need for data-driven solutions to combat air pollution in Poland.

While the proposed methods demonstrate high accuracy in forecasting particulate matter concentrations, it is essential to acknowledge certain limitations. The reliance on data from eight cities in Poland may introduce regional biases, potentially affecting the generalizability of the models to other areas with different pollution sources or environmental conditions. Additionally, the absence of data preprocessing steps, although simplifying the pipeline, may lead to the inclusion of noise or anomalies that could influence the predictions. Another potential limitation concerns the analysis of meteorological data included in the forecasting models. Utilizing meteorological data in air pollution forecasting models is crucial for capturing the complex interactions between weather conditions and pollutant dispersion. Factors such as wind speed and direction play a significant role in transporting particulate matter, potentially spreading pollution over long distances or concentrating it in specific areas. Similarly, temperature and humidity can influence chemical reactions in the atmosphere, affecting the formation and dissipation of pollutants. Precipitation, on the other hand, can aid in removing particulate matter from the air, leading to temporary improvements in air quality. By integrating such meteorological variables, forecasting models could provide more accurate and context-aware predictions, offering valuable insights into how weather patterns exacerbate pollution episodes. This approach would enhance the applicability of the proposed methods, ensuring more comprehensive support for air quality management and public health interventions.

Summary

In this subsection, we have provided a detailed evaluation of the proposed forecasting methods through extensive experimentation. This evaluation encompassed both statistical hypothesis testing and error measure analysis. The results from the Kruskal-Wallis ANOVA, enhanced with post hoc analysis, demonstrated that the forecasts generated by our proposed methods are highly accurate when compared to the original data. This analysis also revealed that traditional algorithms from the literature generally exhibit lower accuracy.

The assessment of error metrics, including MAE and RMSE, further validated the effectiveness of the proposed methods. Across all evaluated scenarios, the proposed methods consistently yielded lower error values compared to the state-of-the-art algorithms, underscoring their better performance.

Overall, the analysis confirms that the proposed methods are reliable and effective for estimating pollutant concentrations, particularly when original data is unavailable. This capability is crucial for scenarios where air quality data might be missing due to technical issues at monitoring stations. Accurate prediction of air pollutant concentrations plays a vital role in air quality management, public health protection, and environmental education.

Conclusion

In this paper, we have dealt with forecasting the concentrations of particulate matter 10 and its four selected components in eight cities in Poland. We have proposed four deep learning-based methods for forecasting, which included xLSTM, KAN, TCN, and VAE. The proposed methods were used on different configurations. Extensive experimental evaluation, combining statistical hypotheses verification and error measures for 400 forecast models, confirmed high prediction accuracy. Moreover, experiments revealed the advantage of the proposed methods over several state-of-the-art algorithms from the literature.

The findings of this study demonstrate the potential of advanced deep learning models to provide accurate and timely forecasts of particulate matter concentrations, which can significantly enhance practical applications in urban planning and public health. These forecasts offer policymakers actionable insights for implementing targeted measures, such as traffic restrictions or temporary industrial limitations, during periods of elevated pollution. Additionally, public health systems can utilize these predictions to issue early warnings, enabling vulnerable populations to take precautionary actions. Furthermore, long-term forecasts generated by the proposed methods can inform urban planning decisions, such as optimizing green spaces, improving city layouts to reduce emissions, and strategically locating pollution monitoring stations. By linking these findings to real-world applications, the study underscores the tangible benefits of data-driven approaches in addressing air quality challenges and safeguarding public health.

In future work, we plan to test the models on datasets from regions with different industrial activities, urban densities, or environmental policies, expecting valuable insights into their robustness and versatility. The experiments may be conducted based on data from different countries, where air pollution is varied. Another promising direction involves integrating real-time data streams, such as meteorological parameters, traffic flow, and industrial activity, to enhance the models' predictive capabilities and responsiveness. Utilizing these dynamic data sources could improve the timeliness and accuracy of forecasts, making them even more effective for real-world applications.

Appendix A

A.1 Normality tests

Table 10 Normality tests for all air pollutants in DsWalbrzWyso air quality monitoring station

	$As(PM_{10})$			$Cd(PM_{10})$			$Ni(PM_{10})$			$Pb(PM_{10})$			PM_{10}		
	h	p	S	h	p	S	h	p	S	h	p	S	h	p	S
original	1	0.00	0.80	1	0.00	0.76	1	0.00	0.53	1	0.00	0.41	1	0.00	0.78
xLSTM 3	1	0.00	0.83	1	0.00	0.80	1	0.00	0.62	1	0.00	0.42	1	0.00	0.78
xLSTM 5	1	0.00	0.83	1	0.00	0.79	1	0.00	0.57	1	0.00	0.48	1	0.00	0.77
xLSTM 7	1	0.00	0.85	1	0.00	0.80	1	0.00	0.61	1	0.00	0.36	1	0.00	0.79
KAN 3	1	0.00	0.82	1	0.00	0.77	1	0.00	0.55	1	0.00	0.37	1	0.00	0.92
KAN 5	1	0.00	0.81	1	0.00	0.77	1	0.00	0.55	1	0.00	0.40	1	0.00	0.91
KAN 7	1	0.00	0.82	1	0.00	0.77	1	0.00	0.54	1	0.00	0.42	1	0.00	0.91
TCN 3	1	0.00	0.83	1	0.00	0.78	1	0.00	0.54	1	0.00	0.35	1	0.00	0.88
TCN 5	1	0.00	0.82	1	0.00	0.77	1	0.00	0.55	1	0.00	0.35	1	0.00	0.90
TCN 7	1	0.00	0.82	1	0.00	0.78	1	0.00	0.54	1	0.00	0.38	1	0.00	0.89
VAE	1	0.00	0.82	1	0.00	0.77	1	0.00	0.54	1	0.00	0.41	1	0.00	0.90
exp. sm	1	0.00	0.95	1	0.00	0.96	1	0.00	0.96	1	0.00	0.47	1	0.00	0.96
exp. (Holt)	1	0.00	0.96	1	0.00	0.97	1	0.00	0.95	1	0.00	0.46	1	0.00	0.96
ARIMA	1	0.00	0.96	1	0.00	0.95	1	0.00	0.97	1	0.00	0.47	1	0.00	0.96
r. forest	1	0.00	0.95	1	0.00	0.95	1	0.00	0.96	1	0.00	0.46	1	0.00	0.95

Table 11 Normality tests for all air pollutants in KpToruDziewu air quality monitoring station

	As(PM_{10})			Cd(PM_{10})			Ni(PM_{10})			Pb(PM_{10})			PM_{10}		
	h	p	S	h	p	S	h	p	S	h	p	S	h	p	S
original	1	0.00	0.63	1	0.00	0.61	1	0.00	0.78	1	0.00	0.91	1	0.00	0.81
xLSTM 3	1	0.00	0.68	1	0.00	0.66	1	0.00	0.81	1	0.00	0.15	1	0.00	0.83
xLSTM 5	1	0.00	0.68	1	0.00	0.71	1	0.00	0.79	0	0.16	0.99	1	0.00	0.82
xLSTM 7	1	0.00	0.70	1	0.00	0.67	1	0.00	0.80	0	0.78	1.00	1	0.00	0.82
KAN 3	1	0.00	0.63	1	0.00	0.63	1	0.00	0.80	1	0.00	0.92	1	0.00	0.94
KAN 5	1	0.00	0.64	1	0.00	0.62	1	0.00	0.80	1	0.00	0.92	1	0.00	0.93
KAN 7	1	0.00	0.63	1	0.00	0.64	1	0.00	0.79	1	0.00	0.93	1	0.00	0.94
TCN 3	1	0.00	0.63	1	0.00	0.64	1	0.00	0.80	1	0.00	0.92	1	0.00	0.90
TCN 5	1	0.00	0.63	1	0.00	0.63	1	0.00	0.80	1	0.00	0.92	1	0.00	0.91
TCN 7	1	0.00	0.63	1	0.00	0.66	1	0.00	0.79	1	0.00	0.92	1	0.00	0.92
VAE	1	0.00	0.63	1	0.00	0.63	1	0.00	0.80	1	0.00	0.93	1	0.00	0.92
exp. sm	1	0.00	0.96	1	0.00	0.95	1	0.00	0.95	1	0.00	0.96	1	0.00	0.95
exp. (Holt)	1	0.00	0.95	1	0.00	0.69	1	0.00	0.81	1	0.00	0.92	1	0.00	0.95
ARIMA	1	0.00	0.94	1	0.00	0.86	1	0.00	0.94	1	0.00	0.91	1	0.00	0.95
r. forest	1	0.00	0.95	1	0.00	0.89	1	0.00	0.95	1	0.00	0.95	1	0.00	0.95

Table 12 Normality tests for all air pollutants in LuZaganKochaMOB air quality monitoring station

	As(PM_{10})			Cd(PM_{10})			Ni(PM_{10})			Pb(PM_{10})			PM_{10}		
	h	p	S	h	p	S	h	p	S	h	p	S	h	p	S
original	1	0.00	0.75	1	0.00	0.57	1	0.00	0.42	1	0.00	0.86	1	0.00	0.88
xLSTM 3	1	0.00	0.76	1	0.00	0.63	1	0.00	0.47	1	0.02	0.99	1	0.00	0.86
xLSTM 5	1	0.00	0.76	1	0.00	0.63	1	0.00	0.46	1	0.01	0.99	1	0.00	0.88
xLSTM 7	1	0.00	0.75	1	0.00	0.65	1	0.00	0.44	1	0.00	0.94	1	0.00	0.86
KAN 3	1	0.00	0.74	1	0.00	0.62	1	0.00	0.45	1	0.00	0.87	1	0.00	0.92
KAN 5	1	0.00	0.76	1	0.00	0.60	1	0.00	0.45	1	0.00	0.87	1	0.00	0.94
KAN 7	1	0.00	0.74	1	0.00	0.64	1	0.00	0.42	1	0.00	0.88	1	0.00	0.93
TCN 3	1	0.00	0.75	1	0.00	0.62	1	0.00	0.45	1	0.00	0.87	1	0.00	0.90
TCN 5	1	0.00	0.76	1	0.00	0.60	1	0.00	0.45	1	0.00	0.86	1	0.00	0.92
TCN 7	1	0.00	0.74	1	0.00	0.64	1	0.00	0.43	1	0.00	0.88	1	0.00	0.90
VAE	1	0.00	0.74	1	0.00	0.61	1	0.00	0.45	1	0.00	0.87	1	0.00	0.91
exp. sm	1	0.00	0.89	1	0.00	0.63	1	0.00	0.90	1	0.00	0.86	1	0.00	0.96
exp. (Holt)	1	0.00	0.93	1	0.00	0.60	1	0.00	0.95	1	0.00	0.88	1	0.00	0.96
ARIMA	1	0.00	0.75	1	0.00	0.57	1	0.00	0.93	1	0.00	0.86	1	0.00	0.95
r. forest	1	0.00	0.73	1	0.00	0.58	1	0.00	0.44	1	0.00	0.84	1	0.00	0.95

Table 13 Normality tests for all air pollutants in MpKrakBulwar air quality monitoring station

	$As(PM_{10})$			$Cd(PM_{10})$			$Ni(PM_{10})$			$Pb(PM_{10})$			PM_{10}		
	h	p	S	h	p	S	h	p	S	h	p	S	h	p	S
original	1	0.00	0.74	1	0.00	0.86	1	0.00	0.96	1	0.00	0.79	1	0.00	0.90
xLSTM 3	1	0.00	0.79	1	0.00	0.88	1	0.00	0.96	1	0.00	0.85	1	0.00	0.89
xLSTM 5	1	0.00	0.77	1	0.00	0.85	1	0.00	0.97	1	0.00	0.78	1	0.00	0.90
xLSTM 7	1	0.00	0.79	1	0.00	0.90	1	0.00	0.97	1	0.00	0.76	1	0.00	0.89
KAN 3	1	0.00	0.75	1	0.00	0.87	1	0.00	0.96	1	0.00	0.81	1	0.00	0.91
KAN 5	1	0.00	0.74	1	0.00	0.86	1	0.00	0.97	1	0.00	0.80	1	0.00	0.90
KAN 7	1	0.00	0.74	1	0.00	0.88	1	0.00	0.96	1	0.00	0.80	1	0.00	0.91
TCN 3	1	0.00	0.75	1	0.00	0.88	1	0.00	0.96	1	0.00	0.80	1	0.00	0.81
TCN 5	1	0.00	0.75	1	0.00	0.87	1	0.00	0.96	1	0.00	0.80	1	0.00	0.84
TCN 7	1	0.00	0.74	1	0.00	0.88	1	0.00	0.96	1	0.00	0.81	1	0.00	0.83
VAE	1	0.00	0.75	1	0.00	0.87	1	0.00	0.96	1	0.00	0.80	1	0.00	0.79
exp. sm	1	0.00	0.94	1	0.00	0.91	1	0.00	0.95	1	0.00	0.80	1	0.00	0.95
exp. (Holt)	1	0.00	0.95	1	0.00	0.94	1	0.00	0.95	1	0.00	0.80	1	0.00	0.95
ARIMA	1	0.00	0.95	1	0.00	0.95	1	0.00	0.95	1	0.00	0.79	1	0.00	0.96
r. forest	1	0.00	0.95	1	0.00	0.96	1	0.00	0.96	1	0.00	0.82	1	0.00	0.96

Table 14 Normality tests for all air pollutants in PkMielPogodn air quality monitoring station

	$As(PM_{10})$			$Cd(PM_{10})$			$Ni(PM_{10})$			$Pb(PM_{10})$			PM_{10}		
	h	p	S	h	p	S	h	p	S	h	p	S	h	p	S
original	1	0.00	0.31	1	0.00	0.86	1	0.00	0.38	1	0.00	0.93	1	0.00	0.89
xLSTM 3	1	0.00	0.45	1	0.00	0.88	1	0.00	0.42	1	0.00	0.98	1	0.00	0.91
xLSTM 5	1	0.00	0.50	1	0.00	0.88	1	0.00	0.43	1	0.00	0.45	1	0.00	0.88
xLSTM 7	1	0.00	0.52	1	0.00	0.89	1	0.00	0.42	1	0.00	0.60	1	0.00	0.92
KAN 3	1	0.00	0.34	1	0.00	0.87	1	0.00	0.39	1	0.00	0.94	1	0.00	0.93
KAN 5	1	0.00	0.32	1	0.00	0.87	1	0.00	0.39	1	0.00	0.94	1	0.00	0.90
KAN 7	1	0.00	0.34	1	0.00	0.88	1	0.00	0.37	1	0.00	0.94	1	0.00	0.92
TCN 3	1	0.00	0.34	1	0.00	0.87	1	0.00	0.40	1	0.00	0.93	1	0.00	0.84
TCN 5	1	0.00	0.32	1	0.00	0.86	1	0.00	0.39	1	0.00	0.94	1	0.00	0.85
TCN 7	1	0.00	0.34	1	0.00	0.88	1	0.00	0.38	1	0.00	0.94	1	0.00	0.85
VAE	1	0.00	0.34	1	0.00	0.87	1	0.00	0.39	1	0.00	0.94	1	0.00	0.82
exp. sm	1	0.00	0.95	1	0.00	0.96	1	0.00	0.47	1	0.00	0.93	1	0.00	0.96
exp. (Holt)	1	0.00	0.38	1	0.00	0.88	1	0.00	0.44	1	0.00	0.93	1	0.00	0.95
ARIMA	1	0.00	0.31	1	0.00	0.86	1	0.00	0.96	1	0.00	0.93	1	0.00	0.90
r. forest	1	0.00	0.33	1	0.00	0.94	1	0.00	0.96	1	0.00	0.93	1	0.00	0.90

Table 15 Normality tests for all air pollutants in SlKatoKossut air quality monitoring station

	$As(PM_{10})$			$Cd(PM_{10})$			$Ni(PM_{10})$			$Pb(PM_{10})$			PM_{10}		
	h	p	S	h	p	S	h	p	S	h	p	S	h	p	S
original	1	0.00	0.84	1	0.00	0.83	1	0.00	0.58	1	0.00	0.93	1	0.00	0.93
xLSTM 3	1	0.00	0.84	1	0.00	0.82	1	0.00	0.65	1	0.00	0.91	1	0.00	0.92
xLSTM 5	1	0.00	0.86	1	0.00	0.84	1	0.00	0.65	1	0.00	0.89	1	0.00	0.93
xLSTM 7	1	0.00	0.86	1	0.00	0.82	1	0.00	0.70	1	0.00	0.90	1	0.00	0.92
KAN 3	1	0.00	0.82	1	0.00	0.82	1	0.00	0.66	1	0.00	0.89	1	0.00	0.89
KAN 5	1	0.00	0.84	1	0.00	0.82	1	0.00	0.64	1	0.00	0.89	1	0.00	0.91
KAN 7	1	0.00	0.83	1	0.00	0.82	1	0.00	0.71	1	0.00	0.89	1	0.00	0.91
TCN 3	1	0.00	0.83	1	0.00	0.84	1	0.00	0.67	1	0.00	0.90	1	0.00	0.85
TCN 5	1	0.00	0.84	1	0.00	0.84	1	0.00	0.64	1	0.00	0.88	1	0.00	0.85
TCN 7	1	0.00	0.83	1	0.00	0.82	1	0.00	0.72	1	0.00	0.89	1	0.00	0.84
VAE	1	0.00	0.82	1	0.00	0.83	1	0.00	0.65	1	0.00	0.89	1	0.00	0.83
exp. sm	1	0.00	0.95	1	0.00	0.88	1	0.00	0.75	1	0.00	0.97	1	0.00	0.96
exp. (Holt)	1	0.00	0.96	1	0.00	0.83	1	0.00	0.76	1	0.00	0.94	1	0.00	0.96
ARIMA	1	0.00	0.84	1	0.00	0.95	1	0.00	0.58	1	0.00	0.93	1	0.00	0.95
r. forest	1	0.00	0.82	1	0.00	0.96	1	0.00	0.57	1	0.00	0.94	1	0.00	0.96

Table 16 Normality tests for all air pollutants in WmPuszczaBor air quality monitoring station

	$As(PM_{10})$			$Cd(PM_{10})$			$Ni(PM_{10})$			$Pb(PM_{10})$			PM_{10}		
	h	p	S	h	p	S	h	p	S	h	p	S	h	p	S
original	1	0.00	0.88	1	0.00	0.89	1	0.00	0.93	1	0.00	0.84	1	0.00	0.84
xLSTM 3	1	0.00	0.86	1	0.00	0.92	1	0.00	0.94	1	0.03	0.99	1	0.00	0.90
xLSTM 5	1	0.00	0.90	1	0.00	0.92	1	0.00	0.94	1	0.00	0.92	1	0.00	0.91
xLSTM 7	1	0.00	0.84	1	0.00	0.92	1	0.00	0.93	1	0.00	0.18	1	0.00	0.89
KAN 3	1	0.00	0.86	1	0.00	0.92	1	0.00	0.94	1	0.00	0.89	1	0.00	0.92
KAN 5	1	0.00	0.87	1	0.00	0.89	1	0.00	0.93	1	0.00	0.88	1	0.00	0.91
KAN 7	1	0.00	0.86	1	0.00	0.92	1	0.00	0.93	1	0.00	0.91	1	0.00	0.94
TCN 3	1	0.00	0.87	1	0.00	0.89	1	0.00	0.94	1	0.00	0.86	1	0.00	0.89
TCN 5	1	0.00	0.88	1	0.00	0.88	1	0.00	0.94	1	0.00	0.87	1	0.00	0.87
TCN 7	1	0.00	0.87	1	0.00	0.90	1	0.00	0.93	1	0.00	0.88	1	0.00	0.89
VAE	1	0.00	0.86	1	0.00	0.91	1	0.00	0.94	1	0.00	0.90	1	0.00	0.89
exp. sm	1	0.00	0.87	1	0.00	0.92	1	0.00	0.93	1	0.00	0.88	1	0.00	0.89
exp. (Holt)	1	0.00	0.93	1	0.00	0.94	1	0.00	0.96	1	0.00	0.96	1	0.00	0.96
ARIMA	1	0.00	0.88	1	0.00	0.90	1	0.00	0.94	1	0.00	0.89	1	0.00	0.93
r. forest	1	0.00	0.88	1	0.00	0.89	1	0.00	0.93	1	0.00	0.84	1	0.00	0.96
	0.00	0.89	1	0.00	0.91	1	0.00	0.93	1	0.00	0.88	1	0.00	0.96	

Table 17 Normality tests for all air pollutants in ZpSzczAndrze air quality monitoring station

	$As(PM_{10})$			$Cd(PM_{10})$			$Ni(PM_{10})$			$Pb(PM_{10})$			PM_{10}		
	h	p	S	h	p	S	h	p	S	h	p	S	h	p	S
original	1	0.00	0.25	1	0.00	0.26	1	0.00	0.88	1	0.00	0.87	1	0.00	0.86
xLSTM 3	1	0.00	0.39	1	0.00	0.52	1	0.00	0.91	0	0.10	0.99	1	0.00	0.86
xLSTM 5	1	0.00	0.39	1	0.00	0.28	1	0.00	0.90	1	0.00	0.90	1	0.00	0.87
xLSTM 7	1	0.00	0.41	1	0.00	0.70	1	0.00	0.90	1	0.00	0.98	1	0.00	0.83
KAN 3	1	0.00	0.27	1	0.00	0.30	1	0.00	0.88	1	0.00	0.89	1	0.00	0.93
KAN 5	1	0.00	0.26	1	0.00	0.28	1	0.00	0.88	1	0.00	0.88	1	0.00	0.93
KAN 7	1	0.00	0.27	1	0.00	0.35	1	0.00	0.87	1	0.00	0.89	1	0.00	0.93
TCN 3	1	0.00	0.27	1	0.00	0.30	1	0.00	0.88	1	0.00	0.88	1	0.00	0.90
TCN 5	1	0.00	0.27	1	0.00	0.28	1	0.00	0.88	1	0.00	0.88	1	0.00	0.90
TCN 7	1	0.00	0.27	1	0.00	0.34	1	0.00	0.87	1	0.00	0.89	1	0.00	0.91
VAE	1	0.00	0.27	1	0.00	0.30	1	0.00	0.89	1	0.00	0.89	1	0.00	0.90
exp. sm	1	0.00	0.86	1	0.00	0.33	1	0.00	0.96	1	0.00	0.91	1	0.00	0.94
exp. (Holt)	1	0.00	0.86	1	0.00	0.36	1	0.00	0.95	1	0.00	0.90	1	0.00	0.95
ARIMA	1	0.00	0.87	1	0.00	0.26	1	0.00	0.88	1	0.00	0.87	1	0.00	0.95
r. forest	1	0.00	0.80	1	0.00	0.28	1	0.00	0.94	1	0.00	0.87	1	0.00	0.94

A. 2 Error measures—preliminaries

Symbol markings:

- min stands for the minimal value of the concentration of the selected air pollutant from the data from GIOŚ;
- max stands for the maximum value of the concentration of the selected air pollutant from the data from GIOŚ;
- A number next to the name of the proposed method denotes the size of the window (e.g., TCN 3 denotes the TCN with a window of size 3; KAN 7 denotes the KAN with a window of size 7);
- mean^1 stands for a mean obtained for all air pollutants for a selected method (horizontal row);
- mean^2 denotes the general mean both for the proposed methods and for state-of-the-art methods (in fact, it is a mean of mean^1);
- exp. sm. denotes exponential smoothing algorithm; exp. (Holt) denotes Holt-Winters exponential smoothing algorithm; ARIMA stands for autoregressive integrated moving average algorithm; r. forest denotes the random forest algorithm.

This information concerns all the tables presented in Sections "A.2.1 Full MAE results" and "A.2.2 Full RMSE results."

A.2.1 Full MAE results

Table 18 MAE values for DsWalbrzWyso air quality monitoring station

	$As(PM_{10})$	$Cd(PM_{10})$	$Ni(PM_{10})$	$Pb(PM_{10})$	PM_{10}	mean[1]	mean[2]
min	0.50	0.10	0.50	0.00	3.90	—	—
max	4.48	1.20	5.75	0.15	115.90	—	—
xLSTM 3	0.45	0.10	0.39	0.01	12.46	2.68	2.68
xLSTM 5	0.31	0.07	0.24	0.01	11.54	2.43	
xLSTM 7	0.43	0.15	0.56	0.01	12.89	2.81	
KAN 3	0.44	0.10	0.40	0.01	10.81	2.35	
KAN 5	0.24	0.05	0.24	0.00	9.74	2.05	
KAN 7	0.44	0.15	0.57	0.01	11.15	2.46	
TCN 3	0.45	0.10	0.41	0.01	10.94	2.38	
TCN 5	0.25	0.05	0.24	0.00	9.82	2.07	
TCN 7	0.45	0.15	0.59	0.01	11.61	2.56	
VAE	0.44	0.10	0.40	0.01	11.37	2.46	
exp. sm	2.69	1.00	1.19	0.15	15.21	4.05	3.84
exp. (Holt)	1.75	1.01	1.11	0.15	15.00	3.80	
ARIMA	1.74	1.01	1.37	0.15	14.81	3.82	
r. forest	1.75	1.01	1.06	0.15	14.46	3.69	

Table 19 MAE values for KpToruDziewu air quality monitoring station

	$As(PM_{10})$	$Cd(PM_{10})$	$Ni(PM_{10})$	$Pb(PM_{10})$	PM_{10}	mean[1]	mean[2]
min	0.50	0.10	0.50	0.00	2.00	—	—
max	2.87	0.42	7.21	0.02	90.76	—	—
xLSTM 3	0.17	0.03	0.78	0.00	10.09	2.22	2.22
xLSTM 5	0.12	0.03	0.51	0.00	9.74	2.08	
xLSTM 7	0.18	0.05	1.12	0.00	10.11	2.30	
KAN 3	0.17	0.03	0.80	0.00	9.58	2.12	
KAN 5	0.09	0.02	0.45	0.00	8.52	1.82	
KAN 7	0.16	0.05	1.16	0.00	10.01	2.28	
TCN 3	0.16	0.04	0.79	0.00	10.27	2.25	
TCN 5	0.09	0.02	0.45	0.00	8.61	1.83	
TCN 7	0.16	0.06	1.16	0.00	10.62	2.40	
VAE	0.17	0.03	0.81	0.00	10.53	2.31	
exp. sm	0.65	0.14	1.35	0.00	10.78	2.58	2.64
exp. (Holt)	0.60	0.01	0.28	0.00	11.50	2.48	
ARIMA	0.54	0.10	1.42	0.00	11.86	2.78	
r. forest	0.48	0.12	1.38	0.00	11.66	2.73	

Table 20 MAE values for LuZaganKochaMOB air quality monitoring station

	$As(PM_{10})$	$Cd(PM_{10})$	$Ni(PM_{10})$	$Pb(PM_{10})$	PM_{10}	mean[1]	mean[2]
min	0.50	0.10	0.50	0.00	2.80	—	—
max	11.47	1.10	8.03	0.20	64.30	—	—
xLSTM 3	1.15	0.07	0.53	0.00	7.22	1.80	1.79
xLSTM 5	0.73	0.04	0.33	0.00	6.95	1.61	
xLSTM 7	1.18	0.11	0.79	0.00	7.42	1.90	
KAN 3	1.18	0.07	0.57	0.00	6.67	1.70	
KAN 5	0.65	0.04	0.34	0.00	5.94	1.39	
KAN 7	1.21	0.11	0.78	0.00	7.08	1.84	
TCN 3	1.19	0.07	0.59	0.00	6.97	1.76	
TCN 5	0.66	0.04	0.33	0.00	5.97	1.40	
TCN 7	1.18	0.11	0.78	0.00	7.21	1.86	
VAE	1.19	0.07	0.57	0.00	6.95	1.76	
exp. sm	2.32	0.06	1.04	0.00	7.56	2.20	2.07
exp. (Holt)	2.23	0.02	1.07	0.00	8.20	2.31	
ARIMA	0.00	0.00	1.03	0.00	7.76	1.76	
r. forest	1.15	0.07	0.51	0.00	8.27	2.00	

Table 21 MAE values for MpKrakBulwar air quality monitoring station

	$As(PM_{10})$	$Cd(PM_{10})$	$Ni(PM_{10})$	$Pb(PM_{10})$	PM_{10}	mean[1]	mean[2]
min	0.50	0.10	0.50	0.00	5.46	—	—
max	3.07	1.29	3.62	0.06	92.19	—	—
xLSTM 3	0.23	0.11	0.37	0.01	13.07	2.76	2.69
xLSTM 5	0.15	0.08	0.26	0.00	12.16	2.53	
xLSTM 7	0.25	0.13	0.52	0.01	13.55	2.89	
KAN 3	0.22	0.11	0.34	0.01	12.05	2.54	
KAN 5	0.12	0.06	0.19	0.00	11.25	2.33	
KAN 7	0.22	0.14	0.52	0.00	12.29	2.64	
TCN 3	0.22	0.11	0.34	0.01	11.91	2.52	
TCN 5	0.12	0.06	0.19	0.00	11.16	2.31	
TCN 7	0.22	0.14	0.52	0.01	12.36	2.65	
VAE	0.22	0.11	0.34	0.01	11.88	2.51	
exp. sm	0.67	0.32	0.49	0.00	14.99	3.30	3.33
exp. (Holt)	0.65	0.30	0.51	0.00	15.12	3.32	
ARIMA	0.63	0.25	0.51	0.00	14.67	3.21	
r. forest	0.64	0.26	0.35	0.01	16.19	3.49	

Table 22 MAE values for PkMielPogodn air quality monitoring station

	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}	mean[1]	mean[2]
min	0.50	0.10	0.50	0.00	8.10	—	—
max	1.67	0.65	4.76	0.02	62.11	—	—
xLSTM 3	0.09	0.05	0.26	0.00	7.52	1.58	1.53
xLSTM 5	0.07	0.03	0.18	0.00	7.57	1.57	
xLSTM 7	0.10	0.08	0.35	0.00	7.66	1.64	
KAN 3	0.08	0.04	0.27	0.00	6.33	1.34	
KAN 5	0.05	0.03	0.14	0.00	6.55	1.35	
KAN 7	0.08	0.08	0.35	0.00	7.05	1.51	
TCN 3	0.08	0.04	0.25	0.00	6.88	1.45	
TCN 5	0.05	0.03	0.14	0.00	6.26	1.30	
TCN 7	0.09	0.08	0.35	0.00	6.67	1.44	
VAE	0.08	0.05	0.27	0.00	7.53	1.58	
exp. sm	0.23	0.15	0.23	0.00	7.92	1.71	1.61
exp. (Holt)	0.03	0.01	0.09	0.00	9.10	1.85	
ARIMA	0.01	0.01	0.53	0.00	5.65	1.24	
r. forest	0.07	0.16	0.54	0.00	7.44	1.64	

Table 23 MAE values for SlKatoKossut air quality monitoring station

	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}	mean[1]	mean[2]
min	0.50	0.10	0.50	0.00	5.60	—	—
max	4.15	3.09	33.84	0.08	87.30	—	—
xLSTM 3	0.36	0.33	2.23	0.01	10.06	2.60	2.91
xLSTM 5	0.23	0.21	1.46	0.01	9.93	2.37	
xLSTM 7	0.37	0.53	3.02	0.02	10.67	2.92	
KAN 3	0.34	0.33	2.39	0.01	10.01	2.62	
KAN 5	0.19	0.19	1.35	0.01	9.01	2.15	
KAN 7	0.34	0.53	3.21	0.02	10.16	2.85	
TCN 3	0.34	0.33	2.35	0.01	10.31	2.67	
TCN 5	0.19	0.19	1.45	0.01	9.03	2.17	
TCN 7	0.34	0.54	3.19	0.02	10.20	2.86	
VAE	0.34	0.34	2.35	0.01	10.15	2.64	
exp. sm	0.51	0.28	6.51	0.01	15.07	4.48	4.13
exp. (Holt)	0.59	0.10	6.70	0.00	15.83	4.65	
ARIMA	0.08	0.53	0.50	0.00	17.01	3.62	
r. forest	0.35	0.54	2.26	0.01	15.75	3.78	

Table 24 MAE values for WmPuszczaBor air quality monitoring station

	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}	mean[1]	mean[2]
min	0.04	0.01	0.04	0.00	3.20	—	—
max	1.03	0.26	1.82	0.01	50.02	—	—
xLSTM 3	0.07	0.03	0.25	0.00	6.07	1.28	1.33
xLSTM 5	0.05	0.02	0.16	0.00	5.71	1.19	
xLSTM 7	0.07	0.04	0.39	0.00	6.66	1.43	
KAN 3	0.07	0.03	0.23	0.00	5.90	1.25	
KAN 5	0.04	0.02	0.13	0.00	5.35	1.11	
KAN 7	0.07	0.05	0.39	0.00	6.59	1.42	
TCN 3	0.07	0.03	0.23	0.00	6.38	1.34	
TCN 5	0.04	0.02	0.13	0.00	5.38	1.11	
TCN 7	0.07	0.05	0.39	0.00	6.96	1.49	
VAE	0.07	0.03	0.23	0.00	6.50	1.36	
exp. sm	0.05	0.02	0.20	0.00	6.36	1.33	1.47
exp. (Holt)	0.02	0.01	0.08	0.00	8.00	1.62	
ARIMA	0.01	0.01	0.05	0.00	7.25	1.47	
r. forest	0.07	0.03	0.23	0.00	7.08	1.48	

Table 25 MAE values for ZpSzczAndrze air quality monitoring station

	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}	mean[1]	mean[2]
min	0.50	0.10	0.50	0.00	3.97	—	—
max	2.17	0.25	5.37	0.01	63.32	—	—
xLSTM 3	0.09	0.01	0.62	0.00	7.33	1.61	1.65
xLSTM 5	0.06	0.00	0.45	0.00	6.69	1.44	
xLSTM 7	0.09	0.02	0.82	0.00	7.45	1.68	
KAN 3	0.07	0.01	0.55	0.00	7.01	1.53	
KAN 5	0.04	0.00	0.32	0.00	5.65	1.20	
KAN 7	0.07	0.01	0.82	0.00	7.28	1.64	
TCN 3	0.07	0.01	0.56	0.00	7.48	1.62	
TCN 5	0.04	0.02	0.32	0.00	5.74	1.22	
TCN 7	0.06	0.01	0.81	0.00	7.80	1.74	
VAE	0.07	0.01	0.57	0.00	7.67	1.67	
exp. sm	0.22	0.01	1.32	0.00	9.35	2.18	2.11
exp. (Holt)	0.22	0.00	1.38	0.00	9.67	2.25	
ARIMA	0.22	0.00	0.14	0.00	9.09	1.89	
r. forest	0.20	0.01	1.47	0.00	8.91	2.12	

A.2.2 Full RMSE results

Table 26 RMSE values for DsWalbrzWyso air quality monitoring station

	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}	mean[1]	mean[2]
min	0.50	0.10	0.50	0.00	3.90	—	—
max	4.48	1.20	5.75	0.15	115.90	—	—
xLSTM 3	0.71	0.17	0.94	0.02	18.67	4.10	3.92
xLSTM 5	0.54	0.13	0.67	0.01	17.54	3.78	
xLSTM 7	0.67	0.22	1.16	0.02	18.97	4.21	
KAN 3	0.72	0.18	1.00	0.02	15.85	3.55	
KAN 5	0.50	0.12	0.71	0.01	15.28	3.33	
KAN 7	0.72	0.22	1.22	0.02	16.27	3.69	
TCN 3	0.72	0.18	1.00	0.02	16.25	3.63	
TCN 5	0.50	0.13	0.71	0.01	15.40	3.35	
TCN 7	0.72	0.22	1.22	0.02	16.70	3.78	
VAE	0.72	0.18	1.00	0.02	16.55	3.69	
exp. sm	2.83	1.05	1.33	0.17	20.28	5.13	4.88
exp. (Holt)	1.90	1.06	1.27	0.17	19.70	4.82	
ARIMA	1.90	1.06	1.49	0.17	19.63	4.85	
r. forest	1.91	1.06	1.22	0.17	19.16	4.70	

Table 27 RMSE values for KpToruDziewu air quality monitoring station

	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}	mean[1]	mean[2]
min	0.50	0.10	0.50	0.00	2.00	—	—
max	2.87	0.42	7.21	0.02	90.76	—	—
xLSTM 3	0.37	0.07	1.42	0.00	15.19	3.41	3.28
xLSTM 5	0.28	0.05	1.05	0.00	15.00	3.28	
xLSTM 7	0.36	0.10	1.78	0.00	14.91	3.43	
KAN 3	0.37	0.07	1.51	0.00	14.13	3.22	
KAN 5	0.27	0.05	1.07	0.00	13.00	2.88	
KAN 7	0.37	0.09	1.85	0.00	14.24	3.31	
TCN 3	0.38	0.07	1.51	0.00	15.00	3.39	
TCN 5	0.27	0.05	1.07	0.00	12.98	2.88	
TCN 7	0.37	0.09	1.85	0.00	14.81	3.43	
VAE	0.38	0.07	1.52	0.00	14.94	3.38	
exp. sm	0.71	0.17	1.70	0.00	14.43	3.40	3.49
exp. (Holt)	0.65	0.04	0.73	0.00	15.30	3.34	
ARIMA	0.59	0.14	1.73	0.00	15.78	3.65	
r. forest	0.53	0.15	1.71	0.00	15.43	3.57	

Table 28 RMSE values for LuZaganKochaMOB air quality monitoring station

	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}	mean[1]	mean[2]
min	0.50	0.10	0.50	0.00	2.80	—	—
max	11.47	1.10	8.03	0.20	64.30	—	—
xLSTM 3	2.03	0.14	1.45	0.00	9.90	2.70	2.67
xLSTM 5	1.49	0.10	1.01	0.00	9.78	2.48	
xLSTM 7	2.14	0.20	1.82	0.00	10.25	2.88	
KAN 3	2.20	0.15	1.54	0.00	8.90	2.56	
KAN 5	1.51	0.10	1.10	0.00	8.36	2.21	
KAN 7	2.20	0.21	1.83	0.00	9.39	2.73	
TCN 3	2.20	0.15	1.55	0.00	9.46	2.67	
TCN 5	1.52	0.10	1.10	0.00	8.56	2.26	
TCN 7	2.20	0.21	1.83	0.00	9.65	2.78	
VAE	2.21	0.15	1.54	0.00	9.23	2.63	
exp. sm	3.04	0.11	1.73	0.00	10.30	3.04	2.88
exp. (Holt)	2.84	0.07	1.72	0.00	10.76	3.08	
ARIMA	0.00	0.00	1.72	0.00	10.66	2.48	
r. forest	2.11	0.14	1.39	0.00	11.02	2.93	

Table 29 RMSE values for MpKrakBulwar air quality monitoring station

	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}	mean[1]	mean[2]
min	0.50	0.10	0.50	0.00	5.46	—	—
max	3.07	1.29	3.62	0.06	92.19	—	—
xLSTM 3	0.38	0.17	0.54	0.01	17.54	3.73	3.58
xLSTM 5	0.30	0.13	0.40	0.01	16.47	3.46	
xLSTM 7	0.41	0.18	0.67	0.01	17.66	3.79	
KAN 3	0.41	0.17	0.53	0.01	16.20	3.47	
KAN 5	0.28	0.12	0.37	0.01	15.14	3.18	
KAN 7	0.41	0.19	0.68	0.01	16.20	3.50	
TCN 3	0.41	0.17	0.54	0.01	16.13	3.45	
TCN 5	0.28	0.12	0.38	0.01	15.13	3.18	
TCN 7	0.41	0.19	0.69	0.01	16.49	3.56	
VAE	0.41	0.17	0.54	0.01	16.06	3.44	
exp. sm	0.81	0.40	0.62	0.01	18.78	4.12	4.09
exp. (Holt)	0.77	0.36	0.64	0.00	18.71	4.10	
ARIMA	0.76	0.31	0.63	0.00	17.85	3.91	
r. forest	0.74	0.31	0.53	0.01	19.65	4.25	

Table 30 RMSE values for PkMielPogodn air quality monitoring station

	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}	mean[1]	mean[2]
min	0.50	0.10	0.50	0.00	8.10	—	—
max	1.67	0.65	4.76	0.02	62.11	—	—
xLSTM 3	0.24	0.08	0.70	0.00	10.15	2.23	2.14
xLSTM 5	0.17	0.06	0.53	0.00	10.13	2.18	
xLSTM 7	0.24	0.12	0.86	0.00	10.24	2.29	
KAN 3	0.25	0.08	0.73	0.00	8.79	1.97	
KAN 5	0.17	0.05	0.51	0.00	8.84	1.92	
KAN 7	0.25	0.12	0.88	0.00	9.38	2.13	
TCN 3	0.25	0.08	0.71	0.00	9.64	2.14	
TCN 5	0.17	0.06	0.50	0.00	8.79	1.90	
TCN 7	0.25	0.13	0.88	0.00	9.44	2.14	
VAE	0.25	0.08	0.73	0.00	9.93	2.20	
exp. sm	0.31	0.18	0.56	0.00	10.02	2.21	2.18
exp. (Holt)	0.11	0.04	0.35	0.00	11.50	2.40	
ARIMA	0.11	0.04	0.89	0.00	8.11	1.83	
r. forest	0.24	0.19	0.89	0.00	10.09	2.28	

Table 31 RMSE values for SlKatoKossut air quality monitoring station

	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}	mean[1]	mean[2]
min	0.50	0.10	0.50	0.00	5.60	—	—
max	4.15	3.09	33.84	0.08	87.30	—	—
xLSTM 3	0.60	0.57	4.91	0.02	13.43	3.90	4.23
xLSTM 5	0.43	0.40	3.82	0.01	13.22	3.58	
xLSTM 7	0.60	0.81	5.89	0.02	14.16	4.30	
KAN 3	0.60	0.58	5.18	0.02	13.42	3.96	
KAN 5	0.44	0.41	3.54	0.01	12.42	3.36	
KAN 7	0.60	0.82	6.09	0.02	13.68	4.24	
TCN 3	0.60	0.58	5.10	0.02	13.94	4.05	
TCN 5	0.43	0.41	3.61	0.01	12.48	3.39	
TCN 7	0.60	0.82	6.06	0.02	13.96	4.29	
VAE	0.60	0.58	5.09	0.02	13.78	4.01	
exp. sm	0.65	0.44	9.33	0.01	18.06	5.70	5.41
exp. (Holt)	0.73	0.28	9.62	0.01	19.20	5.97	
ARIMA	0.30	0.66	2.63	0.01	20.66	4.85	
r. forest	0.58	0.68	5.20	0.02	19.14	5.12	

Table 32 RMSE values for WmPuszczaBor air quality monitoring station

	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}	mean[1]	mean[2]
min	0.04	0.01	0.04	0.00	3.20	—	—
max	1.03	0.26	1.82	0.01	50.02	—	—
xLSTM 3	0.12	0.05	0.39	0.00	8.56	1.82	1.83
xLSTM 5	0.09	0.04	0.28	0.00	7.90	1.66	
xLSTM 7	0.12	0.06	0.54	0.00	9.10	1.96	
KAN 3	0.13	0.05	0.39	0.00	8.35	1.78	
KAN 5	0.09	0.04	0.28	0.00	7.27	1.53	
KAN 7	0.13	0.06	0.54	0.00	8.87	1.92	
TCN 3	0.13	0.05	0.39	0.00	8.76	1.87	
TCN 5	0.09	0.04	0.28	0.00	7.38	1.56	
TCN 7	0.13	0.06	0.54	0.00	9.35	2.02	
VAE	0.13	0.05	0.39	0.00	8.97	1.91	
exp. sm	0.09	0.04	0.31	0.00	8.81	1.85	1.95
exp. (Holt)	0.06	0.03	0.19	0.00	10.13	2.08	
ARIMA	0.06	0.03	0.20	0.00	9.45	1.95	
r. forest	0.12	0.05	0.39	0.00	9.04	1.92	

Table 33 RMSE values for ZpSzczAndrze air quality monitoring station

	As(PM_{10})	Cd(PM_{10})	Ni(PM_{10})	Pb(PM_{10})	PM_{10}	mean[1]	mean[2]
min	0.50	0.10	0.50	0.00	3.97	—	—
max	2.17	0.25	5.37	0.01	63.32	—	—
xLSTM 3	0.24	0.03	1.03	0.00	10.22	2.30	2.31
xLSTM 5	0.17	0.02	0.79	0.00	9.89	2.17	
xLSTM 7	0.23	0.03	1.23	0.00	10.55	2.41	
KAN 3	0.24	0.03	0.99	0.00	9.66	2.18	
KAN 5	0.16	0.02	0.71	0.00	8.44	1.87	
KAN 7	0.24	0.03	1.24	0.00	9.82	2.27	
TCN 3	0.24	0.03	1.00	0.00	10.27	2.31	
TCN 5	0.16	0.02	0.72	0.00	8.60	1.90	
TCN 7	0.24	0.03	1.23	0.00	10.40	2.38	
VAE	0.24	0.03	1.00	0.00	10.38	2.33	
exp. sm	0.32	0.02	1.63	0.00	11.82	2.76	2.69
exp. (Holt)	0.31	0.01	1.70	0.00	12.10	2.82	
ARIMA	0.31	0.01	0.51	0.00	11.45	2.46	
r. forest	0.31	0.03	1.77	0.00	11.50	2.72	

Author Contributions All the research presented in this paper was conducted by its author. The manuscript has been written by its author.

Funding The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability The data used in this research is available at the GIOS website: https://www.gov.pl/web/gios-en

Declarations

Ethical Approval This is an observational study. No ethical approval is required.

Consent to Participate Not applicable—no research involving human subjects was conducted.

Consent for Publication Not applicable—no individuals participated in a study.

Conflict of Interest The author declares no competing interests.

References

Abirami S, Chitra P (2021) Regional air quality forecasting using spatiotemporal deep learning. J Clean Prod 283:125341

Ahmadi MH, Jashnani H, Chau K-W, Kumar R, Rosen MA (2023) Carbon dioxide emissions prediction of five middle eastern countries using artificial neural networks. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects. 45(3):9513–9525

Baklanov A, Zhang Y (2020) Advances in air quality modeling and forecasting. Glob Transit 2:261–270

Beck M, Pöppel K, Spanring M, Auer A, Prudnikova O, Kopp M, Klambauer G, Brandstetter J, Hochreiter S (2024) xLSTM: extended long short-term memory. arXiv:2405.04517

Bekkar A, Hssina B, Douzi S, Douzi K (2021) Air-pollution prediction in smart city, deep learning approach. J Big Data 8:1–21

Bhatti UA, Yan Y, Zhou M, Ali S, Hussain A, Qingsong H, Yu Z, Yuan L (2021) Time series analysis and forecasting of air pollution particulate matter (PM 2.5): an SARIMA and factor analysis approach. IEEE Access 9:41019–41031

Cai B, Yang S, Gao L, Xiang Y (2023) Hybrid variational autoencoder for time series forecasting. Knowl-Based Syst 281:111079

Chae S, Shin J, Kwon S, Lee S, Kang S, Lee D (2021) PM10 and PM2.5 real-time prediction models using an interpolated convolutional neural network. Sci Rep 11(1):11952

Chang Y-S, Chiao H-T, Abimannan S, Huang Y-P, Tsai Y-T, Lin K-M (2020) An LSTM-based aggregated model for air pollution forecasting. Atmos Pollut Res 11(8):1451–1463

Chang Y-S, Abimannan S, Chiao H-T, Lin C-Y, Huang Y-P (2020) An ensemble learning based hybrid model and framework for air pollution forecasting. Environ Sci Pollut Res 27:38155–38168

Cho JH, Moon JW (2022) Integrated artificial neural network prediction model of indoor environmental quality in a school building. J Clean Prod 344:131083

Czernecki B, Marosz M, Jędruszkiewicz J (2021) Assessment of machine learning algorithms in short-term forecasting of PM10 and PM2.5 concentrations in selected polish agglomerations. Aerosol Air Qual Res 21(7):200586. https://doi.org/10.4209/aaqr.200586

Dobrea M, Bădicu A, Barbu M, Subea O, Bălănescu M, Suciu G, Bîrdici A, Orza O, Dobre C (2020) Machine learning algorithms for air pollutants forecasting. In: 2020 IEEE 26th International symposium for design and technology in electronic packaging (SIITME). IEEE, pp 109–113

Doreswamy KS, H., KM Y, Gad I (2020) Forecasting air pollution particulate matter (PM2.5) using machine learning regression models. Procedia Comput Sci **171**, 2057–2066. https://doi.org/10.1016/j.procs.2020.04.221. Third International Conference on Computing and Network Communications (CoCoNet'19)

Du S, Li T, Yang Y, Horng S-J (2019) Deep air quality forecasting using hybrid deep learning framework. IEEE Trans Knowl Data Eng 33(6):2412–2424

Gaj K, Miller U, Sówka I (2020) Progressing climate changes and deteriorating air quality as one of the biggest challenges of sustainable development of the present decade in Poland. MDPI

Gilik A, Ogrenci AS, Ozmen A (2022) Air quality prediction using CNN+ LSTM-based hybrid deep learning architecture. Environ Sci Pollut Res, 1–19

Gu Y, Li B, Meng Q (2022) Hybrid interpretable predictive machine learning model for air pollution prediction. Neurocomputing 468:123–136

Guo Q, He Z, Wang Z (2023a) Predicting of daily pm2 5 concentration employing wavelet artificial neural networks based on meteorological elements in shanghai, China. Toxics 11(1):51

Guo Q, He Z, Wang Z (2023b) Prediction of hourly PM2.5 and PM10 concentrations in Chongqing city in China based on artificial neural network. Aerosol Air Qual Res 23(6):220448. https://doi.org/10.4209/aaqr.220448

Guo Q, He Z, Wang Z (2023c) Simulating daily PM2. 5 concentrations using wavelet analysis and artificial neural network with remote sensing and surface observation data. Chemosphere 340:139886

Guo Q, He Z, Li S, Li X, Meng J, Hou Z, Liu J, Chen Y (2020) Air pollution forecasting using artificial and wavelet neural networks with meteorological conditions. Aerosol Air Qual Res 20(6):1429–1439. https://doi.org/10.4209/aaqr.2020.03.0097

He Z, Guo Q, Wang Z, Li X (2022) Prediction of monthly PM2. 5 concentration in Liaocheng in China employing artificial neural network. Atmosphere 13(8):1221

Heuvelmans MA, van Ooijen PMA, Ather S, Silva CF, Han D, Heussel CP, Hickes W, Kauczor H-U, Novotny P, Peschl H, Rook M, Rubtsov R, von Stackelberg O, Tsakok MT, Arteta C, Declerck J, Kadir T, Pickup L, Gleeson F, Oudkerk M (2021) Lung cancer prediction by deep learning to identify benign lung nodules. Lung Cancer 154:1–4. https://doi.org/10.1016/j.lungcan.2021.01.027

Heydari A, Majidi Nezhad M, Astiaso Garcia D, Keynia F, De Santoli L (2022) Air pollution forecasting application based on deep learning model and optimization algorithm. Clean Technol Environ Policy, 1–15

Hong Zhang PWYQ, Sheng Zhang Wang H (2017) Forecasting of particulate matter time series using wavelet analysis and wavelet-ARMA/ARIMA model in Taiyuan, China. J Air Waste Manag Assoc 67(7), 776–788. https://doi.org/10.1080/10962247.2017.1292968. PMID: 28278031

Ke H, Gong S, He J, Zhang L, Cui B, Wang Y, Mo J, Zhou Y, Zhang H (2022) Development and application of an automated air quality forecasting system based on machine learning. Sci Total Environ 806:151204

Kingma DP, Welling M (2013) Auto-encoding variational Bayes. arXiv:1312.6114

Klambauer G, Unterthiner T, Mayr A, Hochreiter S (2017) Self-normalizing neural networks. CoRR. abs/1706.02515 arXiv:1706.02515

Krajny EA, Ośródka L, Wojtylak MJ (2024) Application of Doppler sodar in short-term forecasting of PM 10 concentration in the air in Krakow (Poland). Atmos Meas Tech 17(8):2451–2464

Kruskal W, Wallis W (1952) Use of ranks in one-criterion variance analysis. J Am Stat Assoc, 583–621

Kruskal WH (1952) A nonparametric test for the several sample problem. Ann Math Stat 23(4):525–540. https://doi.org/10.1214/aoms/1177729332

Kryza M, Werner M, Denby BR, Mu Q, Sawiński T, Remut A (2021) Forecasting PM2. 5 concentrations with uEMEP and eMEP4PL for Poland. In: International technical meeting on air pollution modelling and its application. Springer, pp 193–197

Kryza M, Werner M, Guzikowski J (2020) Assimilation of meteorological data in online integrated atmospheric transport model-example of air quality forecasts for Poland. In: Air pollution modeling and its application XXVI 36. Springer, pp 273–278

Kujawska J, Kulisz M, Oleszczuk P, Cel W (2022) Machine learning methods to forecast the concentration of PM10 in Lublin, Poland. Energies 15(17):6428

Kurnaz G, Demir AS (2022) Prediction of SO2 and PM10 air pollutants using a deep learning-based recurrent neural network: case of industrial city Sakarya. Urban Clim 41:101051

Lea C, Flynn MD, Vidal R, Reiter A, Hager GD (2017) Temporal convolutional networks for action segmentation and detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 156–165

Lea C, Vidal R, Reiter A, Hager GD (2016) Temporal convolutional networks: a unified approach to action segmentation. In: Hua G, Jégou H (eds) Computer Vision - ECCV 2016 Workshops - Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III. Lecture Notes in Computer Science, vol. 9915, pp 47–54. https://doi.org/10.1007/978-3-319-49409-8_7

Lee D, Kang S, Shin J (2017) Using deep learning techniques to forecast environmental consumption level. Sustainability 9(10):1894

Lee M, Lin L, Chen C-Y, Tsao Y, Yao T-H, Fei M-H, Fang S-H (2020) Forecasting air quality in Taiwan by using machine learning. Sci Rep 10(1):4153

Liang Y, Ma J, Tang C, Ke N, Wang D (2023) Hourly forecasting on PM2. 5 concentrations using a deep neural network with meteorology inputs. Environ Monit Assess 195(12):1510

Liao Q, Zhu M, Wu L, Pan X, Tang X, Wang Z (2020) Deep learning for air quality forecasts: a review. Curr Pollut Rep 6:399–409

Liu H, Jin K, Duan Z (2019) Air PM2. 5 concentration multi-step forecasting using a new hybrid modeling method: comparing cases for four cities in China. Atmos Pollut Res 10:1588–1600

Liu Z, Wang Y, Vaidya S, Ruehle F, Halverson J, Soljačić M, Hou TY, Tegmark M (2024) KAN: Kolmogorov-Arnold networks. https://arxiv.org/abs/2404.19756

Liu H, Yin S, Chen C, Duan Z (2020) Data multi-scale decomposition strategies for air pollution forecasting: a comprehensive review. J Clean Prod 277:124023

Liu H, Yan G, Duan Z, Chen C (2021) Intelligent modeling strategies for forecasting air quality time series: a review. Appl Soft Comput 102:106957

Liu B, Yu X, Chen J, Wang Q (2021) Air pollution concentration forecasting based on wavelet transform and combined weighting forecasting model. Atmos Pollut Res 12(8):101144

Majewski G, Szelag B, Mach T, Rogula-Kozlowska W, Aniol E, Bihalowicz J, Dmochowska A, Bihalowicz JS (2021) Predicting the number of days with visibility in a specific range in Warsaw (Poland) based on meteorological and air quality data. Front Environ Sci 9:623094

Mao W, Jiao L, Wang W, Wang J, Tong X, Zhao S (2021) A hybrid integrated deep learning model for predicting various air pollutants. GIScience Remote Sens 58(8):1395–1412

Masood A, Ahmad K (2021) A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: fundamentals, application and performance. J Clean Prod 322:129072

Méndez M, Merayo MG, Núñez M (2023) Machine learning algorithms to forecast air quality: a survey. Artif Intell Rev 56(9):10031–10066

Mishra D, Goyal P (2015) Development of artificial intelligence based NO2 forecasting models at Taj Mahal, Agra. Atmos Pollut Res 6(1):99–106

Murillo-Escobar J, Sepulveda-Suescun J, Correa M, Orrego-Metaute D (2019) Forecasting concentrations of air pollutants using support vector regression improved with particle swarm optimization: case study in Aburrá Valley, Colombia. Urban Clim 29:100473

Nair V, Hinton GE (2010) Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International conference on machine learning (ICML-10), pp 807–814

Nilashi M, Ahmadi H, Manaf AA, Rashid TA, Samad S, Shahmoradi L, Aljojo N, Akbari E (2020) Coronary heart disease diagnosis through self-organizing map and fuzzy support vector machine with incremental updates. Int J Fuzzy Syst 22:1376–1388

Pak U, Ma J, Ryu U, Ryom K, Juhyok U, Pak K, Pak C (2020) Deep learning-based PM2 5 prediction considering the spatiotemporal correlations: a case study of Beijing, China. Sci Total Environ 699:133561

Połednik B (2022) Emissions of air pollution in industrial and rural region in Poland and health impacts. J Ecol Eng 23(9):250–258

Qi Y, Li Q, Karimian H, Liu D (2019) A hybrid model for spatiotemporal forecasting of PM2. 5 based on graph convolutional neural network and long short-term memory. Sci Total Environ 664:1–10

Saxena N (2021) Efficient downscaling of satellite oceanographic data with convolutional neural networks. SIGSPATIAL Special 12(3):46–47

Shams SR, Jahani A, Kalantary S, Moeinaddini M, Khorasani N (2021) The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models for predicting SO2 concentration. Urban Clim 37:100837

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52(3/4):591–611

Sharma M, Jain S, Mittal S, Sheikh TH (2021) Forecasting and prediction of air pollutants concentrates using machine learning techniques: the case of India. In: IOP conference series: materials science and engineering, vol. 1022. IOP Publishing, p 012123

Shi K, Wu L (2020) Forecasting air quality considering the socio-economic development in Xingtai. Sustain Cities Soc 61:102337

Sohn SH, Oh SC, Yeo Y-K (1999) Prediction of air pollutants by using an artificial neural network. Korean J Chem Eng 16:382–387

Song C, Fu X (2020) Research on different weight combination in air quality forecasting models. J Clean Prod 261:121169

Sówka I, Chlebowska-Styś A, Pachurka Ł, Rogula-Kozłowska W, Mathews B (2019) Analysis of particulate matter concentration variability and origin in selected urban areas in Poland. Sustainability 11(20):5735

Subramaniam S, Raju N, Ganesan A, Rajavel N, Chenniappan M, Prakash C, Pramanik A, Basak AK, Dixit S (2022) Artificial intelligence technologies for forecasting air pollution and human health: a narrative review. Sustainability 14(16). https://doi.org/10.3390/su14169951

Udristioiu MT, Mghouchi YE, Yildizhan H (2023) Prediction, modelling, and forecasting of PM and AQI using hybrid machine learning. J Clean Prod 421:138496

Usha Ruby A, Chandran JGC, Theerthagiri P, Patil R, Chaithanya B, Jain TS (2024) Forecasting PM2. 5 concentration using gradient-boosted regression tree with CNN learning model. Opt Mem Neural Netw 33(1):86–96

Usmani RSA, Pillai TR, Hashem IAT, Marjani M, Shaharudin R, Latif MT (2021) Air pollution and cardiorespiratory hospitalization, predictive modeling, and analysis using artificial intelligence techniques. Environ Sci Pollut Res 28(40):56759–56771

Vovk T, Kryza M, Werner M (2024) Using random forest to improve emep4pl model estimates of daily PM2. 5 in Poland. Atmos Environ, 120615

Wang J, Niu T, Wang R (2017) Research and application of an air quality early warning system based on a modified least squares support vector machine and a cloud model. Int J Environ Res Public Health 14(3):249

Yang H, Zhu Z, Li C, Li R (2020) A novel combined forecasting system for air pollutants concentration based on fuzzy theory and optimization of aggregation weight. Appl Soft Comput 87:105972

Yu Y, Li H, Sun S, Li Y (2022) PM2. 5 concentration forecasting through a novel multi-scale ensemble learning approach considering intercity synergy. Sustain Cities Soc 85:104049

Zeinalnezhad M, Chofreh AG, Goni FA, Klemeš JJ (2020) Air pollution prediction using semi-experimental regression model and adaptive neuro-fuzzy inference system. J Clean Prod 261:121218

Zhang L, Lin J, Qiu R, Hu X, Zhang H, Chen Q, Tan H, Lin D, Wang J (2018) Trend analysis and forecast of pm2. 5 in Fuzhou, China using the ARIMA model. Ecol Indic 95:702–710

Zhu D, Cai C, Yang T, Zhou X (2018) A machine learning approach for air quality prediction: model regularization and optimization. Big Data Cogn Comput 2(1):5

A.4 Artykuł A4 – Załącznik 4

Theoretical and Applied Climatology (2025) 156:662
https://doi.org/10.1007/s00704-025-05886-1

RESEARCH

A hybrid transformer and symbolic regression model for weather-dependent particulate matter forecasting in air quality management

Jarosław Bernacki[1]

Received: 29 August 2025 / Accepted: 31 October 2025

Abstract

Forecasting air pollution concentrations, with a particular focus on particulate matter (PM), is a critical component of environmental monitoring systems, enabling timely warnings of smog episodes and the implementation of preventive measures to protect public health. This study introduces a hybrid forecasting approach that combines a transformer-based deep learning model with symbolic regression to predict weather-dependent PM concentrations. The transformer architecture is used to capture complex temporal dependencies in meteorological variables influencing PM variability, while symbolic regression provides interpretable mathematical expressions linking meteorological conditions to PM levels. The proposed method was evaluated using data from multiple air quality monitoring stations, achieving high predictive accuracy while maintaining model transparency, and was compared with several state-of-the-art baseline models. The results demonstrate the efficacy of the proposed approach, achieving an average MAE of approximately 0.10, compared to over 0.20 for the baseline models, an MSE of around 0.02 versus 0.08, and an RMSE of 0.15 compared to 0.25–0.30 for competing methods. These results confirm that the proposed hybrid model delivers substantially more accurate forecasts while preserving interpretability.

Keywords Particulate matter forecasting · Weather-dependent air pollution · Transformer model · Symbolic regression · Environmental monitoring

1 Introduction

Forecasting the concentration of air pollutants, especially particulate matter (PM) particles (PM_{10}, $PM_{2.5}$), is crucial for air quality management and public health protection. Air pollution, especially in cities with high population density and intensive road traffic, has a negative impact on the respiratory and cardiovascular systems, leading to many diseases and premature deaths. Therefore, forecasting PM concentration allows for taking preventive measures, such as issuing warnings about harmful levels of pollution or implementing appropriate emission regulations. Additionally, forecasts enable better planning of health policy and monitoring of the effectiveness of actions aimed at improving air quality. Modern techniques, such as deep learning, offer the possibility of more accurate and personalized forecasting, taking into account dynamic changes in meteorological conditions and other factors influencing the concentration of pollutants. Such an approach is necessary for the effective protection of human health and in achieving the goals of sustainable development in cities Połednik (2022).

Particulate matter (PM) is categorized based on particle size, which determines its aerodynamic behavior and health impact. The main fractions include PM10 (particles smaller than 10 μm), PM2.5 (smaller than 2.5 μm), and PM1 (smaller than 1 μm). Forecasting PM concentrations employs both classical statistical models and advanced deep learning methods. While autoregressive models perform adequately in simple conditions, neural network architectures such as recurrent and convolutional networks are more effective at capturing complex nonlinear dependencies and temporal dynamics in air quality data Abirami (2021).

In this study, we propose a hybrid forecasting model that combines a transformer network with symbolic regression to predict PM concentrations. The transformer

✉ Jarosław Bernacki
jaroslaw.bernacki@pwr.edu.pl

[1] Department of Computer Science and Systems Engineering , Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, Wrocław 50-370, Poland

Published online: 19 November 2025

component captures complex temporal dependencies between meteorological factors, while symbolic regression provides interpretable mathematical expressions linking these variables to pollutant levels. This approach enables both high predictive accuracy and model transparency, offering valuable insights into the underlying mechanisms affecting air quality.

1.1 Contribution

Air pollution forecasting, particularly for PM10 and PM2.5, remains a challenging problem due to the nonlinear and dynamic relationships between meteorological factors and pollutant concentrations. Traditional statistical models and many deep learning architectures either lack interpretability or struggle to generalize under varying environmental conditions. To address these challenges, this paper proposes a hybrid method that combines the transformer architecture with symbolic regression for forecasting PM concentrations. The transformer effectively captures long-range temporal dependencies in meteorological data such as temperature, wind speed, and wind direction, while symbolic regression provides interpretable mathematical expressions linking meteorological variables to PM levels. This integration bridges the gap between high predictive accuracy and model transparency, offering both strong performance and physical interpretability. The novelty of the proposed approach lies in coupling a state-of-the-art deep learning model with an interpretable regression framework, enabling not only precise forecasts but also insights into the environmental mechanisms driving air pollution. The method's effectiveness has been validated on historical datasets, demonstrating its reliability and potential as a foundation for further, more comprehensive analyses.

1.2 Organization of the paper

The next section discusses the previous and related works. In the Sect. 3, the proposed hybrid method is described, in Sect. 4, the experimental evaluation is performed. Finally, the last section concludes this work.

2 Previous and related work

Modern approaches to forecasting air pollution include artificial intelligence techniques, machine learning and statistical methods. For example, convolutional neural networks (CNN) combined with long short-term memory (LSTM) networks Ahmed et al. (2023); Ferrari and Guariso (2024); He et al. (2023); Muruganandam and Arumugam (2022); Zhang et al. (2023) often outperform traditional models. Hybrid methods that combine neural networks, optimization algorithms, and statistical models are gaining popularity due to their ability to model complex data relationships and improve forecast accuracy Du et al. (2022); Kumar et al. (2021); Neto et al. (2021). Statistical methods like ARIMA, SARIMA, and linear regression are popular for predicting air pollutants due to their simplicity and ability to capture patterns and seasonality Aladag (2021); Bhatti et al. (2021).

In the work by Nguyen et al. (2024), a hybrid deep learning model for air quality index (AQI) forecasting, combining machine learning and statistical methods, is presented. The model utilizes ARIMA to capture linear dependencies in the data, followed by a convolutional neural network and an optimized LSTM network to model nonlinear temporal dependencies. Final tuning is performed using the XGBoost model. The obtained results show significant improvement in forecast accuracy compared to traditional methods, confirming the effectiveness of the proposed approach in air quality prediction.

The study presented by Rosca et al. (2025) presents a software for forecasting and predicting the air quality index in urban areas, integrating data on pollutants, meteorological conditions, and road traffic. The authors developed and compared 19 predictive models, including machine learning algorithms, deep learning, and a forecasting model based on structural component analysis. The best results were obtained for the random forest regression model, achieving $R^2 = 99.59\%$ and very low forecast errors. The developed solution was implemented in a web application with an IoT infrastructure and a real-time alert mechanism.

Kulandhainadar & Krishnamoorthy (2025) proposed a novel approach to air quality index forecasting by combining the fuzzy center merge labeling graph (FCMLG) method with the Takagi-Sugeno fuzzy inference system (FIS). The FCMLG approach enables more effective data fusion and uncertainty reduction compared to traditional methods. The model, based on data from Delhi from 2015 to 2021, integrates information on pollutants and meteorological parameters, achieving high accuracy with low error values. The developed approach, combining regression and classification, achieved 99% prediction accuracy, confirming its effectiveness in air quality modeling.

In the paper proposed by Ordenshiya and Revathi (2024), a new air quality index forecasting method based on a combination of a fuzzy center merge graph with an optimal fuzzy inference system (FCMG-OP-FIS) and machine learning techniques is presented. The proposed model effectively transforms regression data into classification data, integrating pollutant and meteorological data to improve forecast accuracy. The FCMG-OP-FIS regression results are then classified using IF-THEN rules and a random forest classifier model, which achieves 99% accuracy, precision, and recall. The developed approach demonstrates high performance in data labeling and AQI prediction, confirming its usefulness in urban environmental monitoring and management.

The study presented by Kumar et al. (2025) evaluated the performance of various machine learning models in predicting the air quality index for Gurugram (India), with respect to their relevance to achieving the Sustainable Development Goals (SDGs). Decision tree techniques such as M5P, random forest, random tree, and reduced error pruned tree were analyzed using daily data from the Central Pollution Control Board. The results confirm that the M5P-T5 model outperforms the other methods, while the random tree model was the least effective in predicting AQI.

The work by Ansari and Quaff (2025) concerns the prediction of the hourly air quality index in Azamgarh, Uttar Pradesh, using deep learning models. The dataset included 8,760 measurements of particulate matter ($PM_{2.5}$, PM_{10}), gases (NO_2, SO_2), and meteorological parameters. Statistical analyses (MANOVA, ANOVA, *t*-tests) confirmed significant changes in pollutants across timescales and day-night differences. Sensitivity analysis indicated that $PM_{2.5}$, NO_2, and SO_2 have the greatest impact on the forecasts, making the feedforward neural network an effective tool for accurate AQI forecasting.

The study by Ramentol et al. (2023) proposes machine learning-based prediction of nitrogen dioxide NO_2 concentrations, using historical data and meteorological data from the city of Erfurt (Germany). The temporal dependencies through embedded variables, enabling the model to account for local factors such as traffic and special events, are utilized. Additionally, the approach considers pollutant seasonality and meteorological parameters to predict NO_2. Evaluation demonstrates higher forecasting accuracy compared to other models.

An air quality forecasting method based on an attention temporal graph convolutional network (AT-GCN), which combines attention, a recurrent GRU network, and graph convolutional mechanisms, is proposed in Iskandaryan et al. (2023). The method is tested on air quality, meteorological, and road traffic data from Madrid. Results revealed that the proposed approach better captures complex dependencies and achieves lower prediction errors, compared to CNN or LSTM networks.

3 The proposed method

In this section, we describe the proposed hybrid approach utilizing a transformer and symbolic regression for weather-dependent PM concentration forecasting. The transformer is used for the input data pre-processing. It analyzes temporal relations and detects dependencies between samples of the input data. The symbolic regression processes the data returned by the transformer in order to find an optimal mathematical equation that best describes them. In our case, the proposed hybrid approach processes the meteorological data, which includes the temperature, wind speed, and wind direction, in order to predict the PM concentration.

3.1 The transformer

The transformer acts as a feature extractor that analyzes the relationships between different samples in the sequence and creates a more informative representation of them. Thanks to the self-attention mechanism, the model can focus on the important parts of the data, ignoring noise and irrelevant information. For example, if the input data has some seasonality, the transformer detects similarities between distant samples and uses them for prediction. The transformer consists of the embedding layer and the attention mechanism, which we briefly describe below.

3.2 The embedding layer

The embedding layer transforms the input data into a higher-dimensional space, making it easier for the transformer to process it. This allows it to better capture hidden patterns and relationships between input features. The embeddings also play a role in matching the dimensions of the data to the requirements of the transformer architecture. This allows the model to more effectively analyze complex temporal or structural relationships in the data.

3.3 The attention mechanism

The attention mechanism in the transformer allows each element of a sequence to “pay attention” to other elements and adjust its representation based on their importance. The self-attention engine calculates how strongly each data sample influences other samples. Self-attention layers are applied to the input data. The transformer calculates the similarity between each sample and the others, assigning higher weights to those that are more important for future prediction. The weighted values are then summed, creating a new representation of the input, enriched with

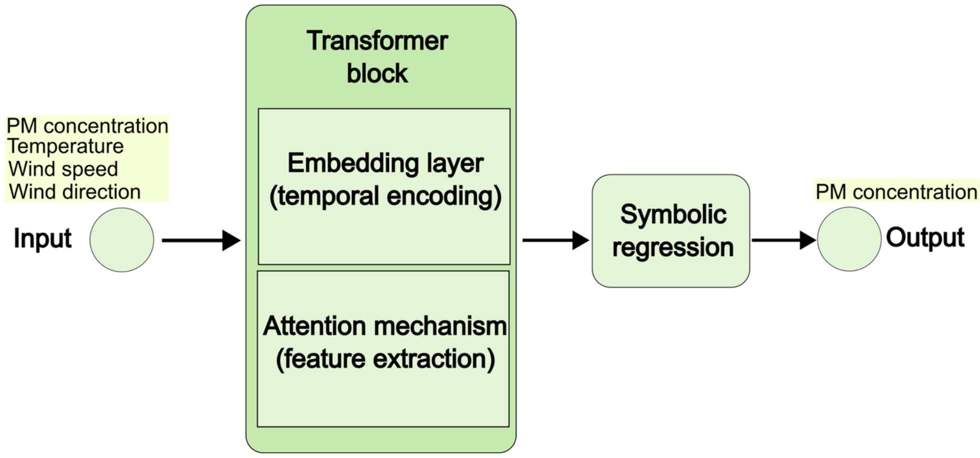


Fig. 1 The structure of the proposed hybrid method

information from the entire sequence. This allows the model to, for example, detect recurring patterns or long-term dependencies in the data. This process is done in parallel for all elements, allowing for efficient analysis of complex structures. As a result, self-attention allows the transformer to better understand the context of the data, which improves the quality of prediction. The linear layer is used to transform the transformer output back to the original data dimension.

3.4 Symbolic regression

Symbolic regression is a method for finding mathematical equations that best describe data. Unlike classical linear or nonlinear regression, there is no preconceived notion of a specific form of the function (e.g., a polynomial of a specific degree or an exponential function). Instead, the algorithm itself builds and tests various mathematical expressions until it finds one that best fits the data.

Symbolic regression explores the space of equations composed of basic mathematical operators. Mathematical formulas can be represented as expression trees, where each node is a mathematical operation (e.g., addition, multiplication) and the leaves are variables or constants. The algorithm starts with simple expressions (e.g., random combinations of variables and basic operations) and then modifies them through evolutionary operations, such as mutation (replacing one function with another, e.g., x^2 with $\sin(x)$), crossover (combining parts of different expressions), and adding new operators.

Each equation is evaluated based on a fitness function, e.g. mean square error (MSE) between predicted values and actual data. Better formulas are promoted to the next generation. The process is iterative – formulas evolve in successive generations, and less accurate ones are eliminated. The goal is to find the simplest and most accurate function. Therefore, this process is similar to the training of the neural network.

The idea of the proposed method is summarized in the Fig. 1.

4 Experimental evaluation

In this section, we benchmark the proposed hybrid forecasting method using real-world data from the Polish Chief Inspectorate of Environmental Protection (GIOS)[1]The dataset consists of hourly measurements of particulate matter (PM_{10} and $PM_{2.5}$) concentrations, expressed in µg/m³, covering the period from January 1 to December 31, 2022.

To account for meteorological influences, we include weather variables obtained from the Meteostat[2] service for the same time interval. Specifically, temperature (°C), wind speed (km/h), and wind direction (degrees) were selected due to their completeness and their established impact on air pollution variability.

The proposed model was trained on the meteorological data and used to forecast PM_{10} and $PM_{2.5}$ concentrations, which were subsequently validated against the observed GIOS measurements. Data were normalized to the [0,1] range prior to training.

The implementation was carried out in Python using the PyTorch framework with CUDA acceleration, while symbolic regression was performed with the PySR library. Statistical evaluation of the results was conducted in MATLAB. All computations were performed on a notebook equipped with an Intel Core i7-13700 H processor, 32 GB of RAM, and an NVIDIA RTX 4070 GPU.

[1] https://www.gov.pl/web/gios-en.

[2] https://meteostat.net/en/.

4.1 Collected data and area description

The forecasting evaluation was conducted using PM_{10} and $PM_{2.5}$ concentrations from selected air quality monitoring stations in Poland: KpZielBoryTu (Zielonka Bory Tucholskie), MpZakopaSien (Zakopane), SlCzestoBacz (PM_{10})/SlCzestoZana ($PM_{2.5}$) (Częstochowa), SlKnurJedNar (PM_{10})/SlGliwicMewy ($PM_{2.5}$) (Gliwice/Knurów), and ZpSzczec1Maj (Szczecin). These codes, with the city names in parentheses, are used throughout the paper. The stations were selected based on data completeness. Although in some cities PM_{10} and $PM_{2.5}$ are measured at different nearby stations (Częstochowa: SlCzestoBacz/SlCzestoZana; Gliwice/Knurów: SlKnurJedNar/SlGliwicMewy), their close proximity ensures that this does not negatively affect the evaluation. The locations of the air quality monitoring stations are shown in Fig. 2.

4.2 Hyperparameters

All hyperparameters are determined experimentally. The transformer's embedding dimension (vector representation of the input data) was set to 64, with 4 attention heads and 16 transformer layers. The model was trained using the Adam optimizer with a learning rate of 0.0001 and the mean squared error (MSE) as the loss function, for 500 epochs. For symbolic regression, 1000 iterations are used to identify the mathematical equation.

To reduce the risk of overfitting, the model was trained and evaluated on separate, non-overlapping datasets. The obtained results were consistent across different monitoring stations and time periods, confirming the robustness and generalization capability of the proposed approach.

4.3 The symbolic regression equation

The procedure of the symbolic regression allowed to find the following equation for calculating the predicted concentration of the PM y:

$$y = b_1 \cdot x_1^2 + b_2 \cdot \sin(b_3 x_2) + b_4 \cdot \cos(b_5 x_3)$$

The coefficients b_i ($i = 1,\ldots,5$), also determined by the symbolic regression method, were equal: b_1 = -0.23, b_2 = -0.75,

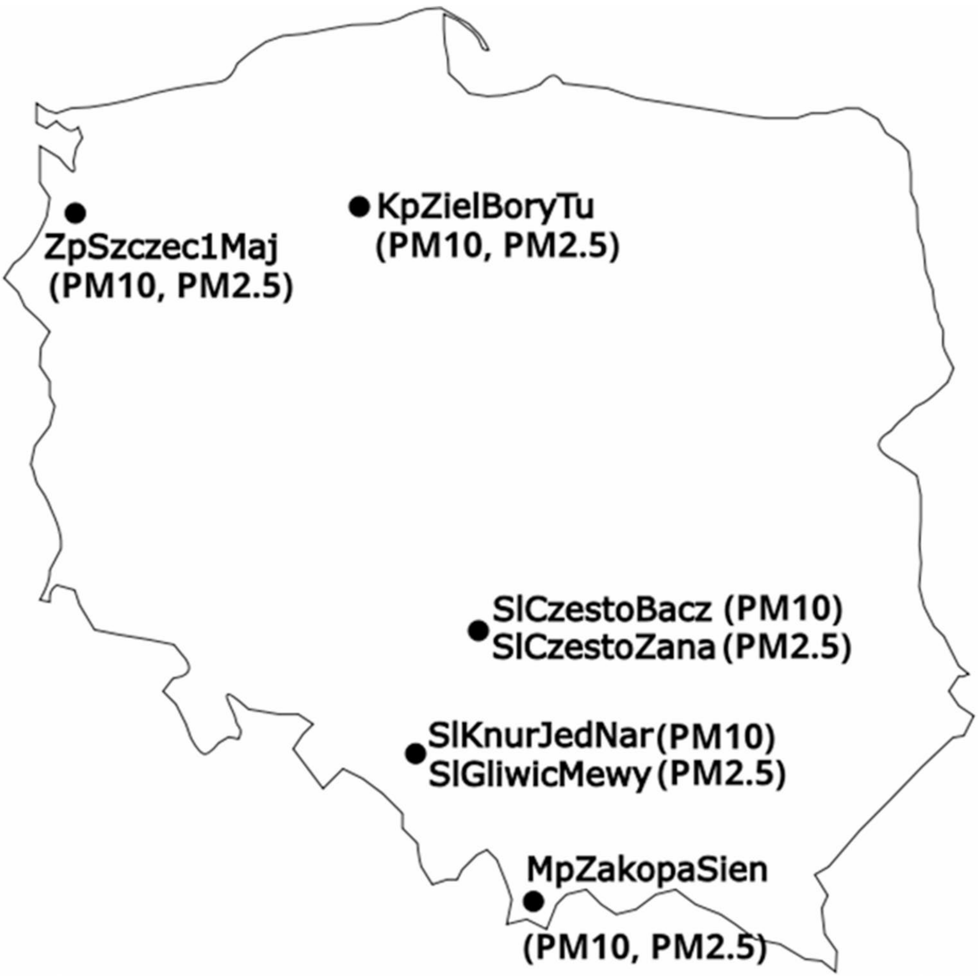


Fig. 2 The location of the air quality monitoring stations in Poland used in the experiment

$b_3 = 1.12$, $b_4 = 0.42$, $b_5 = 0.70$. The values x_1, x_2, and x_3 correspond to the following meteorological data: temperature, wind speed, and wind direction, respectively. The coefficients in the symbolic regression equation were obtained through an optimization process to minimize prediction error, rather than being derived from physical principles. Therefore, they do not have a direct physical interpretation. However, the functional forms and included meteorological variables provide insight into the relationships affecting PM concentrations.

Example 1 Let us evaluate the equation with the following data from January 04 (2022) at the station KpZielBoryTu. The normalized values for the temperature x_1, wind speed x_2, and wind direction x_3 are as follows: $x_1 = 0.40$, $x_2 = 0.39$, and $x_3 = 0.65$. We will use these values to estimate the concentration of the PM_{10}:

$$y = -0.23 \cdot (0.40)^2 + (-0.75) \cdot \sin(1.12 \cdot 0.39) + 0.42 \cdot \cos(0.70 \cdot 0.65) = 0.19$$

The real (historical) concentration of the PM_{10} was on that day equal to 0.20 (normalized). This confirms the almost perfect predicting accuracy of the proposed method.

Example 2 Let us evaluate the equation with the following data from May 27 (2022) at station ZpSzczec1Maj. The normalized values for the temperature x_1, wind speed x_2, and wind direction x_3 are as follows: $x_1 = 0.58$, $x_2 = 0.41$, and $x_3 = 0.72$. We will use these values to estimate the concentration of the $PM_{2.5}$:

$$y = -0.23 \cdot (0.58)^2 + (-0.75) \cdot \sin(1.12 \cdot 0.41) + 0.42 \cdot \cos(0.70 \cdot 0.72) = 0.11.$$

The real (historical) concentration of the $PM_{2.5}$ was on that day equal to 0.11 (normalized). This confirms the perfect predicting accuracy of the proposed method.

4.4 Baseline models – state-of-the-art methods

4.4.1 Multilayer perceptron

A multilayer perceptron (MLP) is a classical artificial neural network composed of several fully connected (linear) layers with nonlinear activation functions. The structure and main features of the MLP, following Gardner (1998), are as follows:

- The first linear layer transforms the input data into 64 neurons;
- The second linear layer maps data from 64 to 128 neurons;
- The third linear layer reduces the data back from 128 to 64 neurons;
- The final layer produces the desired output using a 1×1 convolution, generating a prediction tensor for the sequence.

The network was trained for 1000 epochs using the Adam optimizer with a learning rate of 0.001. Each layer uses the ReLU (rectified linear unit) activation function, defined as $\mathrm{ReLU}(x) = \max(0, x)$, where x is the input to the neuron. All hyperparameters were determined experimentally. The graphical structure of the network is shown in Fig. 3.

The MLP is trained using historical data consisting of PM concentrations, temperature, wind speed, and wind direction. These data were organized into tuples, each containing a single value of the four variables at a given time step t. The tuples served as input to the MLP, with only the PM concentration values used as labels. This setup defines the objective function, allowing the model to generate PM concentration forecasts by minimizing the error between the predicted and historical PM concentrations. The predicted PM concentration at a single time step is denoted as y_t.

4.4.2 Recurrent neural network

A recurrent neural network (RNN) is a network that has the ability to process sequences of data by updating a hidden state at each time step Rumelhart (1986). The proposed RNN is a model designed to analyze sequential data such as time series. The model uses an architecture consisting of an RNN layer with 64 units (cells), which allows it to capture

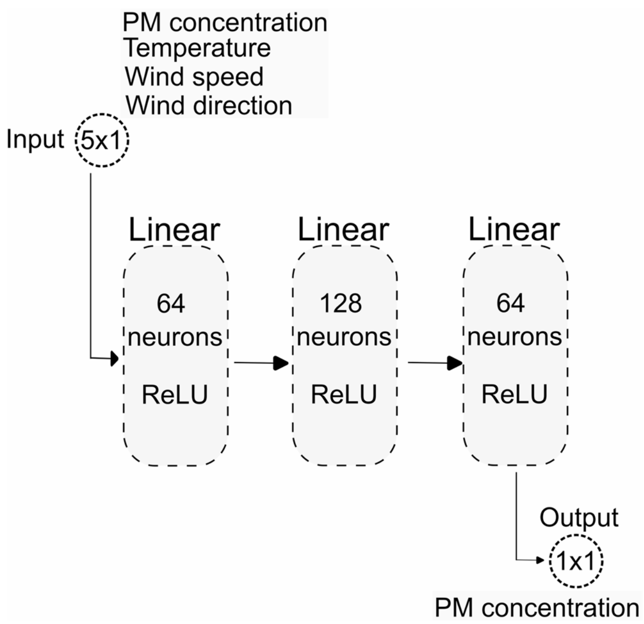


Fig. 3 The architecture of multilayer perceptron base model

complex temporal relationships in the data. The units are connected to each other.

The network processes the input data in the form of a time sequence by RNN layers, which analyze it in the context of previous time steps. In each iteration, the network maintains a hidden state that stores information about the previous data using a hyperbolic tangent function. This mechanism allows the model to efficiently model the temporal relationships in the data, which is crucial for forecasting future values.

More formally, RNN operations can be defined as follows. The hidden state h_t is updated based on the input x_t, and the previous hidden state h_{t-1} in the following manner:

$$ht = \tan h(W_x X_t + W_h h_{t-1} + b_h)$$

Where:

- W_x is a weight matrix connecting input x_t with the actual state;
- W_h is a weight matrix connecting the previous state, h_{t-1}, with the actual state;
- b_h is a bias vector;
- tanh is a hyperbolic tangent function.

Graphically, the RNN cell is visualized in Fig. 4.

The output of the last RNN layer is a representation of the data from the last time step. This layer transforms the hidden representation into a single numerical value – a forecast of the future concentration of the PM. The model is optimized using the Adam algorithm with a mean squared error (MSE) loss function minimizing the difference between the predicted and actual values.

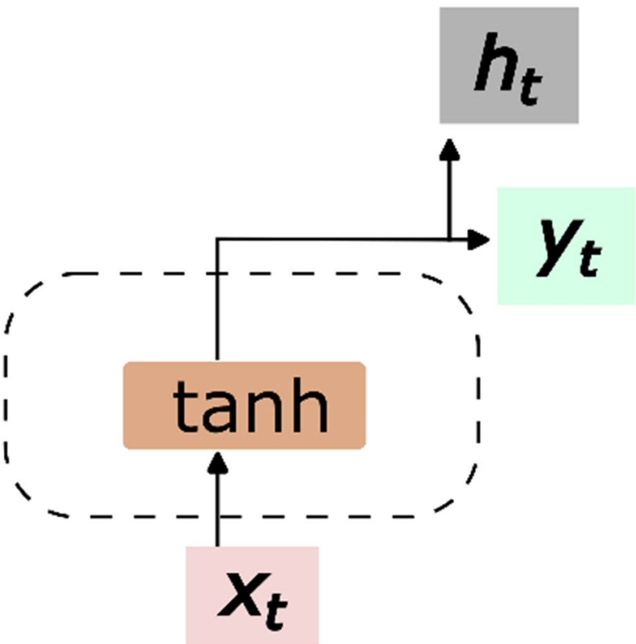


Fig. 4 The structure of the RNN cell

4.4.3 ARIMAX

ARIMA (autoregressive integrated moving average) is a model for analyzing and forecasting time series. It consists of three components: the autoregressive part, the integrated part which stands for differencing the time series to achieve stationarity, and the moving average part, which takes into account errors from previous forecasts. The algorithm takes as input the three values: (p,d,q). The p denotes the order of autoregression (the number of previous values of the target variable that are used); the d is the number of differences needed to achieve stationarity; and q is the order of moving average (the number of previous errors that are used). The ARIMAX is an extension of ARIMA, which enables using external variables in the predictions Williams (2001).

In our experiments, we have used the following parameters: $p=5$, which denotes taking 5 last observations for the prediction; $d=1$ enables the differentiation only once; $q=1$ states that the algorithm takes into account the impact of the forecast error from the previous period t. The prediction calculation using the ARIMAX method is described below:

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \Theta_j \varepsilon_{t-j} + \beta^T X_t + \varepsilon_t$$

Where:

- y_t – the predicted value of particulate matter concentration;
- ϕ – the impact of the previous values in the time series;
- θ – the impact of the previous prediction errors;
- $\beta^T x_t$ – the product of a scalar vector of coefficients β and a vector x_t of the external variables.
- $\beta = [\beta_1, \beta_2, \beta_3, \beta_4]$.
- $x_t = [x^{conc}, x^{temp}, x^{wind}, x^{wdir}, x^{pres}]$.
- ε – a random noise at time t;
- p – the number of elements in x_t.

The external variables are represented as x^{temp}, x^{wind}, x^{wdir}, x^{pres}, and denote respectively the temperature, wind speed, wind direction, and air pressure. The weights ϕ, θ, and β are estimated using MLE (maximum likelihood estimation).

4.4.4 k-Nearest neighbors regression

k-Nearest Neighbors Regression (k-NNR) is a simple method based on the similarity of observations Fix and

Hodges (1951). It assumes that observations with similar values of input features (e.g., weather conditions) also have similar values of the dependent variable (e.g., PM concentration). The forecast is obtained by analyzing the *k* most similar cases from the past. The idea of *k*-NNR operations is summarized as follows:

1. For each forecasted weather situation, the distance (e.g., Euclidean) between current conditions and all historical data is calculated;
2. The *k* observations that are most similar are selected;
3. The forecast result is the average PM value among these *k* neighbors.

In our case, for each point in time *t*, we create an input feature vector:

$$X_t = [x^{conc},\ x^{temp},\ x^{wind},\ x^{wdir},\ x^{pres}]$$

Where x^{conc}, x^{temp}, x^{wind}, x^{wdir}, x^{pres} denote respectively the historical PM concentration, temperature, wind speed, wind direction, and air pressure. Next, we create the following dataset: (X_t, y_t) where $y_t = x_{t+1}^{conc}$. For a new time step t' with current weather conditions:

1. We calculate the distance (e.g., Euclidean) between $X_{t'}$ and all training points X_t;
2. We select *k* nearest neighbors;
3. Finally, we calculate the average of their corresponding y_t, which is the predicted PM concentration in a time step $t+1$.

We have evaluated the algorithm by using the data from $t-5$ former time steps.

4.4.5 Linear regression with time delays

Linear regression (LR) is a classical statistical technique for modeling the relationship between a dependent variable and a set of predictors Galton (1886). In the context of PM forecasting, the model can be extended with so-called "lags" – historical PM values from previous time intervals in order to better capture the dynamics of the phenomenon. The operations of the algorithm can be summarized as follows:

1. An input feature vector is created containing current meteorological data and historical PM concentrations;
2. The model estimates linear coefficients assigned to each variable;
3. The predicted PM value is a weighted sum of the input variables and the error term.

In our case, we create the following feature vector with delays:

$$X_t = [X_t^{conc},\ X_{t-1}^{conc},\ \ldots\ X_t^{temp},\ X_{t-1}^{temp},\ \ldots\ X_t^{wind},\ X_{t-1}^{wind},\ \ldots\ X_t^{wdir}, X_{t-1}^{wdir},\ \ldots\ X_t^{pres},\ X_{t-1}^{pres}, \ldots]$$

The target variable is $y_t = x_{t+1}^{conc}$. Then, we train the linear regression on data (X_t, y_t) to let the model fit linear coefficients. To obtain the forecast of the next PM value, we put the current weather data and the latest PM values into the model. Similarly, as in the case of *k*-NNR, we use the data from the five $(t-5)$ former time steps.

4.5 Evaluation results

To evaluate the performance of the forecasts generated by the proposed method, we calculate the following error metrics: mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE), where x_t is the actual value, y_t is a predicted value, and n is the number of observations, as defined below.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - x_t}{n}\right|$$

$$MSE = \frac{1}{n}\sum\nolimits_{t=1}^{n}(x_t - y_t)^2$$

$$RMSE = \sqrt{\frac{1}{n}\sum\nolimits_{t=1}^{n}(x_t - y_t)^2}$$

In Tables 1 and 2 there are presented results of error measures for predicting the concentration values of PM_{10} and $PM_{2.5}$ for all air quality monitoring stations.

The analysis of prediction error results for all measurement stations clearly indicates that the proposed method (SR) outperforms the other approaches, achieving the lowest MAE, MSE, and RMSE values in all cases. At each of the stations – regardless of the location and data characteristics – the SR method shows significantly lower errors compared to methods based on neural networks (MLP, RNN), classical statistical models (ARIMAX, LR), and *k*-NNR. For example, for the MpZakopaSien station, the SR method achieves an RMSE of 0.11, while the competitive approaches are characterized by values exceeding 0.20. Also, at the other stations, these differences are significant, reaching even 0.15 in RMSE values, which indicates a clear advantage of SR in terms of forecast accuracy. Importantly, SR both minimizes mean absolute errors, but also limits the occurrence of large deviations of predictions from actual values, as evidenced by lower MSE values. The obtained results confirm the effectiveness and stability of the proposed approach, making it a promising tool in applications related to air pollution forecasting.

Table 1 Comparison of error metrics (MAE, MSE, and RMSE) between the predicted and actual PM_{10} concentrations for the proposed symbolic regression (SR) method and baseline models

Air station	Measure	SR	MLP	RNN	ARIMAX	k-NNR	LR
	MAE	0.13	0.23	0.21	0.22	0.21	0.22
KpZielBoryTu	MSE	0.03	0.08	0.07	0.08	0.07	0.07
	RMSE	0.16	0.28	0.26	0.27	0.27	0.27
	MAE	0.08	0.17	0.17	0.20	0.20	0.20
MpZakopaSien	MSE	0.01	0.05	0.04	0.06	0.06	0.06
	RMSE	0.11	0.21	0.21	0.25	0.25	0.25
	MAE	0.11	0.22	0.22	0.22	0.21	0.22
SlCzestoBacz	MSE	0.02	0.07	0.08	0.07	0.07	0.07
	RMSE	0.15	0.27	0.28	0.27	0.26	0.27
	MAE	0.10	0.22	0.20	0.22	0.22	0.22
SlKnurJedNar	MSE	0.02	0.08	0.06	0.07	0.07	0.08
	RMSE	0.14	0.28	0.24	0.27	0.27	0.28
	MAE	0.12	0.19	0.19	0.23	0.22	0.23
ZpSzczec1Maj	MSE	0.03	0.06	0.06	0.08	0.07	0.08
	RMSE	0.17	0.24	0.24	0.28	0.27	0.28

Table 2 Comparison of error metrics (MAE, MSE, and RMSE) between the predicted and actual $PM_{2.5}$ concentrations for the proposed symbolic regression (SR) method and baseline models

Air station	Measure	SR	MLP	RNN	ARIMAX	k-NNR	LR
	MAE	0.11	0.19	0.19	0.22	0.21	0.21
KpZielBoryTu	MSE	0.02	0.06	0.06	0.08	0.07	0.07
	RMSE	0.15	0.24	0.24	0.28	0.27	0.27
	MAE	0.07	0.17	0.19	0.20	0.19	0.20
MpZakopaSien	MSE	0.01	0.05	0.06	0.07	0.06	0.07
	RMSE	0.11	0.21	0.25	0.26	0.25	0.26
	MAE	0.12	0.22	0.20	0.23	0.21	0.21
SlCzestoZana	MSE	0.03	0.08	0.06	0.08	0.07	0.07
	RMSE	0.16	0.28	0.25	0.28	0.27	0.27
	MAE	0.13	0.25	0.26	0.24	0.25	0.24
SlGliwicMewy	MSE	0.03	0.10	0.11	0.09	0.09	0.09
	RMSE	0.17	0.32	0.33	0.31	0.30	0.31
	MAE	0.10	0.22	0.21	0.20	0.21	0.20
ZpSzczec1Maj	MSE	0.02	0.07	0.07	0.07	0.06	0.07
	RMSE	0.14	0.27	0.26	0.26	0.25	0.25

4.6 Visual inspection of the forecasts

In the Figs. 5, 6 and 7 there are presented average MAE, MSE, and RMSE values both from the proposed method and the state-of-the-art baseline methods, respectively.

The proposed hybrid model demonstrated a clear advantage over the baseline forecasting approaches. It achieved an average MAE of approximately 0.10, compared to values exceeding 0.20 for the baseline models, indicating a substantial reduction in average prediction error. Similarly, the MSE reached around 0.02, while competing methods yielded approximately 0.08, confirming that the proposed model more effectively minimizes large deviations. The RMSE further supports these findings, with values of 0.15 for the proposed approach versus 0.25–0.30 for the reference models. These improvements are consistent across both $PM_{2.5}$ and PM_{10} concentrations, demonstrating the robustness and stability of the hybrid transformer–symbolic regression framework. Overall, the results highlight the higher predictive capability of the proposed method while maintaining interpretability through the symbolic regression component.

4.7 Statistical hypotheses verification

We have analyzed at a significance level of $\alpha = 0.05$, if the forecasts generated with the proposed method differ significantly from the original, historical PM concentrations, and also from state-of-the-art methods. The typical procedure of hypothesis testing begins with the normality test, which determines the compliance with the normal distribution. If

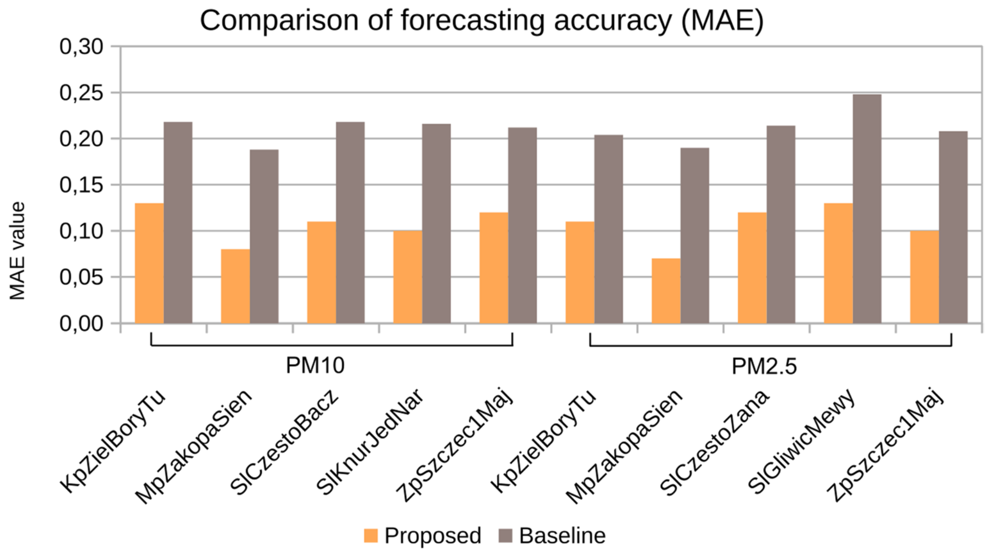


Fig. 5 Comparison of forecasting accuracy: average MAE values between the proposed method and the baseline models across stations

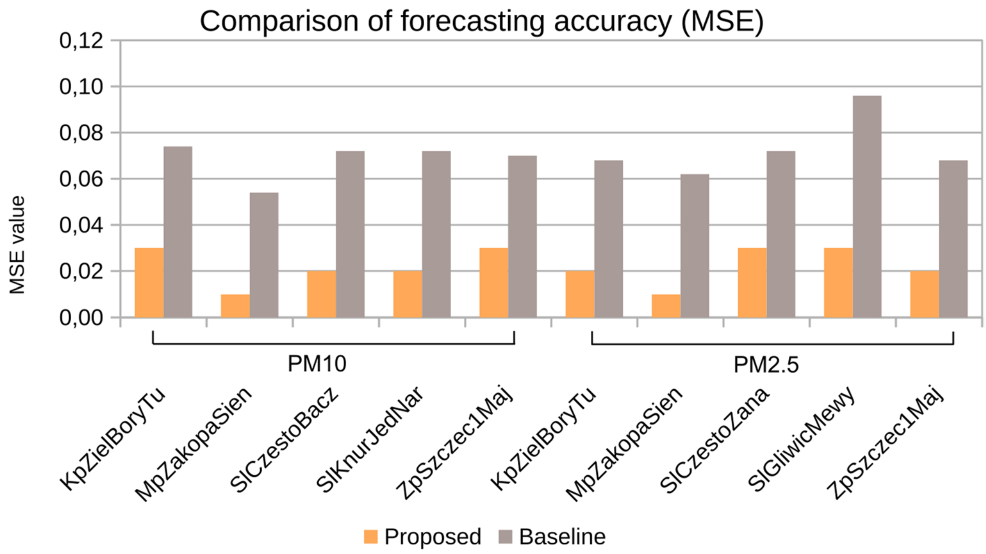


Fig. 6 Comparison of forecasting accuracy: average MSE values between the proposed method and the baseline models across stations

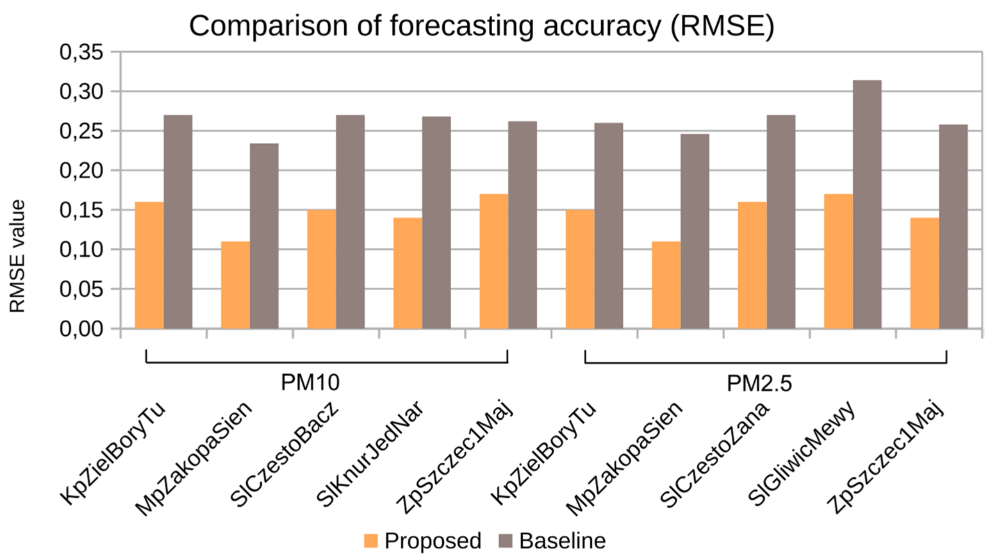


Fig. 7 Comparison of forecasting accuracy: average RMSE values between the proposed method and the baseline models across stations

the analyzed data perform a normal distribution, the parametric tests for further testing may be used, including the variance or average tests. However, the incompatibility with the normal distribution denotes that non-parametric testing should be realized.

4.7.1 Normality tests

The statistical verification begins with the normality tests, which check whether the analyzed data follows the normal distribution. We propose to use the Shapiro-Wilk test. The hypotheses are defined as follows (Shapiro-Wilk, (1965):

- H_0: The analyzed data follows the standard distribution;
- H_1: The analyzed data does not follow the standard distribution.

To accept or reject the null hypothesis, we calculate the Shapiro-Wilk test statistic S and then calculate the p-value (let us further name it shortly p). The p is compared with the significance level, and if greater, then there is no reason to reject the null hypothesis H_0. Otherwise, the alternative hypothesis H_1 is accepted:

- If $p \leq \alpha$: the H_0 is rejected and the H_1 is accepted;
- If $p > \alpha$: there is no reason to reject the H_0.

Alternatively, if S is greater than the test critical value, the H_0 might be accepted. In case of our results, all the p-values are close to 0.00, so the null hypothesis should be rejected. Therefore, we assume that the data does not follow a normal distribution. For further analysis, the non-parametric Kruskal-Wallis ANOVA test should be used.

4.7.2 Kruskal-Wallis ANOVA

The non-parametric Kruskal-Wallis ANOVA (also called Kruskal-Wallis one-way analysis of variance for ranks) is an extension of the U Mann-Whitney test for more than two populations. This test is used for the verification of the insignificance of differences of the medians in $k > 2$ populations Kruskal and Wallis (1952). The hypotheses for k groups are defined as follows:

- H_0: $\Theta_1 = \Theta_2 = \ldots = \Theta_k$ (all the medians are equal);
- H_1: Not all medians Θ_j are equal ($j = 1, 2, \ldots, k$).

To accept or reject the null hypothesis, it is essential to calculate the test statistic H, and then calculate the p. The p is compared with the significance level, and if greater, then there is no reason to reject the null hypothesis H_0. Otherwise, the alternative hypothesis H_1 is accepted:

- If $p \leq \alpha$: the H_0 is rejected and the H_1 is accepted;
- If $p > \alpha$: there is no reason to reject the H_0.

Alternatively, we can compare the test statistic H with the test critical value. If the test statistic is greater than the critical value, the H_0 should be rejected.

In our case, we statistically compare the medians of the original data and the forecasts generated by the proposed methods. Since the total number of compared methods is 6, the number of degrees of freedom between groups is equal to $6–1 = 5$. The number of degrees of freedom within groups is equal to $6 \cdot 365–6 = 2184$. Therefore, the χ^2 critical value is equal to 18.55. Based on the test statistic, we calculate the p. To accept the H_0, the value of the test statistic should not exceed the critical value. Results of the Kruskal-Wallis ANOVA p-values are presented in Table 3 for PM_{10} and $PM_{2.5}$ forecasting.

The results of the Kruskal-Wallis test (ANOVA) indicate that for all the analyzed measurement stations, the H statistic values significantly exceed the critical value, and the accompanying p values are equal to 0.00. This means that there are statistically significant differences between the analyzed data groups. For both the PM_{10} and $PM_{2.5}$ fractions, the null hypothesis of the homogeneity of distributions can therefore be rejected, and it can be assumed that at least one of the compared prediction methods differs significantly from the others.

4.7.3 Post-hoc analysis

Due to the fact that Kruskal-Wallis ANOVA indicates in general, whether the groups differ (or not) significantly without pointing out precise differences, the post-hoc analysis may be performed. The post-hoc indicates the sum of the ranks of the particular groups, so it makes it possible to deduce the tendencies between them. In our case, if the sum of the ranks of a particular method is close to the original

Table 3 Values of H and p of the Kruskal-Wallis ANOVA test – forecasts comparison with original data and state-of-the-art methods for PM_{10} and $PM_{2.5}$

PM fraction	Air station	H	p
PM_{10}	KpZielBoryTu	21.21	0.00
	MpZakopaSien	20.14	0.00
	SlCzestoBacz	37.24	0.00
	SlKnurJedNar	18.75	0.00
	ZpSzczec1Maj	21.21	0.00
$PM_{2.5}$	KpZielBoryTu	35.74	0.00
	MpZakopaSien	21.26	0.00
	SlCzestoZana	33.54	0.00
	SlGliwicMewy	27.13	0.00
	ZpSzczec1Maj	19.64	0.00

Table 4 Results of the post-hoc analysis

PM fraction	Air station	SR	MLP	RNN	ARIMAX	k-NNR	LR
	KpZielBoryTu	1529	1590	1589	1545	1502	1588
	MpZakopaSien	1495	1387	1389	1539	1589	1595
PM_{10}	SlCzestoBacz	1562	1548	1593	1548	1505	1590
	SlKnurJedNar	1645	1334	1694	1430	1476	1435
	ZpSzczec1Maj	1576	1552	1564	1517	1472	1566
	KpZielBoryTu	1453	1566	1611	1630	1584	1537
	MpZakopaSien	1559	1399	1404	1342	1453	1630
$PM_{2.5}$	SlCzestoZana	1644	1311	1597	1663	1426	1574
	SlGliwicMewy	1650	1642	1599	1462	1419	1376
	ZpSzczec1Maj	1564	1438	1436	1490	1538	1483

data, it means that the forecast is accurate. However, if the sum of ranks is higher or lower than the original data, it means that the forecast values are too high or too low. In Table 4 the values of the sum of ranks for the particular data are described.

The analysis of the average ranks indicates that the values assigned to the SR method are well balanced with respect to the actual values. In cases where the average rank of another method is lower than the rank of SR (e.g., MLP and RNN for the SlCzestoZana station in the case of $PM_{2.5}$), it can be concluded that the given method tends to underestimate the forecasted values. Conversely, if the average rank of a given method is higher than the rank of SR (e.g., ARIMAX for most stations), it indicates a systematic overestimation of the results. Such an observation confirms that SR is characterized by greater stability. This is especially visible in the example of the MpZakopaSien station for PM_{10}, where both MLP, RNN, and LR significantly outperform SR in terms of ranks, which suggests a tendency for these methods to generate too high values. At the same time, methods such as MLP and RNN for some stations (e.g., SlKnurJedNar) may underestimate the forecasts, which additionally indicates their instability. In the context of all stations and both particulate matter fractions, the SR method maintains a relatively consistent level, which strengthens its assessment as a reliable and precise method.

4.8 Discussion and limitations

The proposed hybrid forecasting model demonstrates several notable strengths. Combining the transformer architecture with symbolic regression makes the method achieving both high predictive accuracy and interpretability. The approach effectively captures nonlinear temporal dependencies among meteorological variables while producing explicit analytical expressions that describe their influence on PM concentrations. These features make the model both accurate and transparent.

Despite the promising performance of the proposed model, several limitations should be acknowledged. First, the model has been trained and validated using data from specific monitoring stations in Poland, which may limit its generalizability to other geographical areas or different meteorological conditions. Extending the model to multiple locations and assessing its cross-regional performance would be an important direction for future research.

Another limitation concerns the coverage of the dataset. Although the data used in this study are complete and continuous over the 2022 year, they are limited to the selected stations and environmental conditions. Expanding the dataset to include more locations, longer time periods, or diverse meteorological scenarios could further improve model robustness and predictive accuracy.

There is also potential for improving the model itself. While the hybrid approach utilizes the strengths of both deep learning and symbolic regression, future work could explore the usage of additional environmental variables, experimentation with alternative transformer architectures or attention mechanisms, and refinement of the symbolic regression component to generate more interpretable or physically meaningful equations.

Finally, the practical deployment of the model has not yet been evaluated. Investigating its implementation in real-time air quality monitoring systems, with attention to computational efficiency and prediction reliability, would be a valuable step toward operational use. Addressing these limitations could enhance the applicability, interpretability, and overall reliability of the hybrid model for PM forecasting.

4.9 Summary

The experiments carried out showed that the proposed hybrid method achieved the highest efficiency in forecasting particulate matter concentrations. The obtained results were characterized by the lowest error values (MAE, MSE, RMSE) compared to other methods described in the literature. The model better reproduced the temporal and seasonal variability of PM concentrations, which confirms its high precision. The obtained results indicate that the proposed approach is a significant improvement in the prediction of air quality.

Statistical analysis using the Kruskal-Wallis ANOVA test confirmed significant differences between the accuracy of forecasts obtained by different methods. The proposed hybrid model showed statistically significantly better efficiency than the other approaches. Other methods were characterized by higher values of prediction errors, which indicates their lower accuracy. These results confirm the higher quality of forecasts generated by the proposed model and its advantage over solutions from the literature.

5 Conclusion

In this paper, a novel approach to PM concentration forecasting was presented, using a hybrid model combining a transformer and symbolic regression. Integration of meteorological data with transformer-based and symbolic regression models allows for capturing both complex temporal relationships and identifying interpretable patterns that can help in making decisions related to air quality management. Preliminary experimental results show that the proposed method effectively predicts the PM concentration using meteorological data, compared to the baseline models. Future work will concern the further development of the proposed method for a comprehensive experimental study, as well as the improvement of the forecasting accuracy.

Author contributions The research has been conducted and the manuscript was written by its author – Jarosław Bernacki.

Funding Declaration.
The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability The data used in this research is available at the GIOS and Meteostat websites: https://www.gov.pl/web/gios-enhttps://meteostat.net/en/.

Declarations

Competing interests The authors declare no competing interests.

References

Abirami S, Chitra P (2021) Regional air quality forecasting using spatiotemporal deep learning. J Clean Prod 283:125341

Ahmed AN, Ean LW, Chow MF, Malek MA, Zaini N (2023) Forecasting of fine particulate matter based on LSTM and optimization algorithm. J Clean Prod 427:139233

Aladag E (2021) Forecasting of particulate matter with a hybrid ARIMA model based on wavelet transformation and seasonal adjustment. Urban Clim 39:100930

Ansari A, Quaff AR (2025) Data-driven analysis and predictive modelling of hourly air quality index (AQI) using deep learning techniques: a case study of Azamgarh, India. Theor Appl Climatol 156:74. https://doi.org/10.1007/s00704-024-05304-y

Bhatti UA, Yan Y, Zhou M, Ali S, Hussain A, Qingsong H, Yu Z, Yuan L (2021) Time series analysis and forecasting of air pollution particulate matter (pm 2.5): an Sarima and factor analysis approach. IEEE Access 9:41019–41031

Du P, Wang J, Yang W, Niu T (2022) A novel hybrid fine particulate matter (pm2.5) forecasting and its further application system: case studies in China. J Forecast 41:6485

Ferrari L, Guariso G (2024) Convolutional forecasting of particulate matter: toward a data-driven generalized model. Atmosphere 15:398

Fix E, Hodges JL (1951) Discriminatory analysis, nonparametric discrimination

Galton F (1886) Regression towards mediocrity in hereditary stature. The Journal of the Anthropological Institute of Great Britain and Ireland 15:246263

Gardner MW, Dorling S (1998) Articial neural networks (the multilayer perceptron) a review of applications in the atmospheric sciences. Atmos Environ 32(14–15):2627–2636

He J, Li X, Chen Z, Mai W, Zhang C, Wan X, Wang X, Huang M (2023) A hybrid clstm-gpr model for forecasting particulate matter (pm2.5). Atmos Pollut Res 14:101832

Iskandaryan D, Ramos F, Trilles S (2023) Graph neural network for air quality prediction: a case study in Madrid. IEEE Access 11:2729–2742. https://doi.org/10.1109/ACCESS.2023.3234214

Kruskal W, Wallis W (1952) Use of ranks in one-criterion variance analysis. J Am Stat Assoc. https://doi.org/10.1080/01621459.1952.10483441

Kulandhainadar Mariavalavan O, Krishnamoorthy G, R (2025) Enhanced air quality index prediction using fuzzy center merge labeling graph based fuzzy inference system model. Cybernetics Syst 1–43. https://doi.org/10.1080/01969722.2025.2522873

Kumar A, Sihag S, Singh K, Esmaeilbeiki F, Singh, Balraj, Om H (2025) Prediction of air quality index (AQI) using machine learning-based tree models—a case study of gurugram City, India. 156. Theoretical and Applied Climatology10.1007/s00704-025-05535-7

Kumar R, Bhardwaj P, Pfister G, Drews C, Honomichl S, Attilo D, G (2021) Description and evaluation of the fine particulate matter forecasts in the Ncar regional air quality forecasting system. Atmosphere 12:302

Muruganandam NS, Arumugam U (2022) Seminal stacked long short-term memory (ss-lstm) model for forecasting particulate matter (pm2.5 and pm10). Atmosphere 13:1726

Neto PSDM, Firmino PRA, Siqueira H, Tadano YDS, Alves TA, De Oliveira JFL, Marinho MHDN, Madeiro F (2021) Neural-based ensembles for particulate matter forecasting. IEEE Access 9:1447014490

Nguyen AT, Pham DH, Oo B, Ahn Y, Lim BTH (2024) Predicting air quality index using attention hybrid deep learning and quantum-inspired particle swarm optimization. J Big Data 11:71. https://doi.org/10.1186/s40537-024-00926-5

Ordenshiya KM, Revathi GK (2024) Hybrid FCMG-OP-FIS model approach to convert regression into classification data for machine learning-based AQI prediction. Heliyon. https://doi.org/10.1016/j.heliyon.2024.e39759

Połednik B (2022) Emissions of air pollution in industrial and rural region in Poland and health impacts. J Ecol Eng 23:250–258

Ramentol E, Grimm S, Stinzendorfer M, Wagner A (2023) Short-Term Air Pollution Forecasting Using Embeddings in Neural Networks. Atmosphere 14:298

Rosca C-M, Carbureanu M, Stancu A (2025) Data-driven approaches for predicting and forecasting air quality in urban areas. Appl Sci 15(8):4390. https://doi.org/10.3390/app15084390

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back propagating errors. Nature 323(6088):533–536

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52(3/4):591–611

Williams BM (2001) Multivariate vehicular Tracffic flow prediction: evaluation of Arimax modeling. Transp Res Rec 1776(1):194200

Zhang Y, Sun Q, Liu J, Petrosian O (2023) Long-term forecasting of air pollution particulate matter (pm2. 5) and analysis of influencing factors. Sustainability 16:19

A.5 Artykuł A5 – Załącznik 5

DOI: 10.17951/b.2025.80.0.265-283

ANNALES
UNIVERSITATIS MARIAE CURIE-SKŁODOWSKA
LUBLIN – POLONIA

VOL. LXXX SECTIO B 2025

JAROSŁAW BERNACKI
https://orcid.org/0000-0002-4488-3488
Department of Computer Science and Systems Engineering
Wrocław University of Science and Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
jaroslaw.bernacki@pwr.edu.pl

Particulate Matter Forecasting Using Hybrid Autoencoders: The Role of Meteorological Data

Modelowanie i prognozowanie pyłu zawieszonego z wykorzystaniem hybrydowych autoenkoderów i danych meteorologicznych

Abstract: Accurate forecasting of particulate matter (PM) concentrations is crucial for effective air quality management and public health protection. In this study, we propose two deep learning-based forecasting models: a convolutional-recurrent autoencoder and a hierarchical autoencoder. Both models are trained and evaluated using historical data on $PM_{2.5}$ and PM_{10} concentrations from multiple monitoring stations in Poland. To assess the influence of meteorological conditions on prediction accuracy, the models are tested in two variants: one using historical PM concentrations and meteorological features such as temperature, wind speed, wind direction, and air pressure, and another that uses only historical PM data. The results clearly show that the inclusion of weather data significantly improves forecasting performance, with lower MAE, and MSE values observed across all test sites. The models trained with meteorological inputs consistently outperform their counterparts trained on PM data alone. The results are also compared with the baseline model. These findings highlight the importance of environmental context in air pollution forecasting and demonstrate the potential of autoencoder-based approaches for this task.

Keywords: air pollution; forecasting; deep models; particulate matter

Abstrakt: Prognozowanie stężeń pyłu zawieszonego ma kluczowe znaczenie dla efektywnego zarządzania jakością powietrza i ochrony zdrowia publicznego. W opisanym badaniu proponujemy dwa modele prognostyczne oparte na uczeniu głębokim: autokoder konwolucyjno-rekurencyjny oraz autokoder hierarchiczny. Oba modele są trenowane i oceniane z wykorzystaniem danych historycznych dotyczących stężeń pyłu $PM_{2.5}$ i PM_{10} z kilku stacji monitorujących jakość powietrza w Polsce. Aby ocenić wpływ warunków meteorologicznych na dokładność prognoz, modele testowano w dwóch wariantach – jeden wykorzystuje historyczne stężenia pyłu zawieszonego oraz dane meteorologiczne, takie jak temperatura, prędkość i kierunek wiatru oraz ciśnienie powietrza, a drugi wyłącznie historyczne dane dotyczące pyłu zawieszonego. Wyniki pokazują, że uwzględnienie danych pogodowych znacząco poprawia skuteczność prognozowania, z niższymi wartościami MAE i MSE zaobserwowanymi we wszystkich lokalizacjach testowych. Modele trenowane z wykorzystaniem danych meteorologicznych konsekwentnie przewyższają swoje odpowiedniki trenowane wyłącznie na danych dotyczących pyłu zawieszonego. Uzyskane wyniki zostały porównane z wynikami modelu bazowego. Obserwacje te podkreślają znaczenie kontekstu środowiskowego w prognozowaniu zanieczyszczenia powietrza oraz pokazują potencjał podejść opartych na autokoderach.

Słowa kluczowe: zanieczyszczenie powietrza; prognozowanie; modele głębokie; pył zawieszony

INTRODUCTION

Particulate matter (PM) is an air pollutant that consists of small solid and liquid particles suspended in the atmosphere. The most commonly distinguished are the PM_1 with particles of a diameter less than 1 micrometer, $PM_{2.5}$ with particles with a diameter of less than 2.5 micrometers, and PM_{10} with particles with a diameter of less than 10 micrometers. These microscale particles can penetrate deep into the lungs, and in the case of PM_1 and $PM_{2.5}$, also into the bloodstream, which poses a serious health risk, leading to respiratory and cardiovascular diseases and cancer. Due to its harmful health effects, PM is one of the most important environmental problems in the world. Many countries, especially those with a high level of urbanization and industrial development, are trying to reduce PM emissions by introducing strict regulations, developing air purification technologies, and monitoring air quality. Effective forecasting of PM concentration is crucial for taking action to prevent its negative impact on public health and improving the quality of life of city dwellers (Polichetti et al., 2009).

In Poland, PM, especially $PM_{2.5}$ and PM_{10}, is a serious environmental and health problem. Its high concentrations in the air occur mainly during the heating season when intensive home heating contributes to the emission of pollutants from chimneys, as well as a result of road transport and industry. According to reports by the WHO and national institutions, Poland is one of the countries with the highest level of air pollution in the European Union, which negatively affects the health of citizens, especially in cities with high population density (Jasiński et al., 2021). Health effects, such as respiratory and cardiovascular diseases and an increased number of premature deaths, are directly related to exposure to PM. In response to this problem, Poland has taken a number of actions, including the introduction of air quality standards, the modernization of heating systems, and the development of an air quality monitoring network. Despite this, effective reduction of PM emissions remains one of the main challenges for the country's environmental protection policy. The $PM_{2.5}$ and PM_{10} (but also other air pollutants) concentrations are monitored on an ongoing basis by the Chief Inspectorate of Environmental Protection (GIOS)[1], which provides data online (Penkała et al., 2023, Połednik, 2022).

Technical challenges at air quality monitoring stations often result in missing data. To tackle this problem, an effective strategy is to estimate the air pollutants' missing values, which is a task closely tied to forecasting. A variety of studies have been conducted on forecasting air pollutant concentrations (Sowka et al., 2019), utilizing diverse methodologies such as machine learning (Gryech et al., 2024), deep learning (Swetha et al., 2024), and hybrid models (Borah et al., 2024).

In this study, we focus on forecasting the concentrations of $PM_{2.5}$ and PM_{10} using advanced deep learning techniques. Specifically, convolutional-recurrent autoencoders and hierarchical autoencoders models are employed to predict future PM levels based on historical data, including meteorological variables such as temperature, wind speed, wind direction, and air pressure. The choice of these models is motivated by their ability to capture complex nonlinear patterns in air quality data, which traditional methods may struggle with. To evaluate the performance of these models, commonly used metrics such as mean absolute error (MAE), and mean squared error (MSE) are used, ensuring a comprehensive assessment of prediction accuracy. The experiments are conducted on the historical data collected from the following Polish cities: Zielonka (Bory Tucholskie), Zakopane, Częstochowa, Gliwice, and Szczecin.

The forecasting simulations are evaluated twofold: one using historical PM concentrations and meteorological features such as temperature, wind speed, wind direction, and air pressure, and another that uses only historical PM data. The use of meteorological data alongside historical PM concentrations leads to significantly improved forecasting results. Models trained with weather variables consistently achieved lower prediction error values compared to those using only PM historical concentrations. This demonstrates that atmospheric factors play a key role in shaping air pollution dynamics, and their inclusion enhances the models' ability to capture complex temporal patterns and variability in PM levels. In addition, we implemented a state-of-the-art random forest model and tested it analogously to the proposed autoencoders, comparing its performance with and without meteorological inputs under the same experimental setup.

[1] https://www.gov.pl/web/gios-en

CONTRIBUTION

The main contribution of this study lies in the application of advanced deep learning models for forecasting PM concentrations, specifically $PM_{2.5}$ and PM_{10}, using meteorological data. This research evaluates the effectiveness of convolutional-recurrent and hierarchical autoencoder architectures in capturing complex nonlinear dependencies in air pollution data. Meteorological variables are crucial in this context, as weather conditions strongly influence pollutant dispersion, transport, and accumulation. For example, wind speed and direction can disperse pollutants or carry them across regions, significantly altering local concentration levels. By comparing models trained with and without meteorological inputs, the study highlights the added value of utilizing atmospheric variables in prediction tasks. A thorough evaluation using standard performance metrics such as MAE and MSE on historical data demonstrates the potential of the proposed approach to improve the accuracy of air quality forecasts and support environmental and public health decision-making.

ORGANIZATION OF THE PAPER

The paper is organized as follows. The next section describes previous and related works. Section 3 presents proposed deep learning methods for forecasting the concentration of $PM_{2.5}$ and PM_{10}. In section 4, the experimental evaluation of the proposed methods is performed. Finally, the last section concludes this work. Everywhere in the paper, bold font in formal descriptions denotes matrices or vectors.

PREVIOUS AND RELATED WORK

In Li et al. (2020), a hybrid CNN-LSTM model combining a convolutional neural network (CNN) and an LSTM neural network is developed to forecast the $PM_{2.5}$ concentration for the next 24 hours in Beijing. This model takes advantage of both networks: CNN effectively extracts air quality-related features, and LSTM takes into account the long-term temporal context of the data. The model is fed with the air quality data of the past 7 days, and the output is the forecast of the next day's $PM_{2.5}$ concentration. This study compares four models: univariate LSTM, multivariate LSTM, univariate CNN-LSTM, and multivariate CNN-LSTM. The results show that the proposed multivariate CNN-LSTM model achieves the best performance with low error and short training time.

The study presented by Mauricio et al. (2024) investigates the use of a transformer-based model for forecasting coarse particulate matter (PMCO) concentrations in Mexico City over several horizons. The transformer outperformed ARIMA and LSTM models across multiple stations, though challenges remain for short-term forecasts and areas with high variability near industrial zones.

The work depicted by Harishkumar et al. (2020) determines the concentration of PM in the atmosphere to improve public health. In this study, machine learning models were used to predict the concentration of particulate matter based on data from the air quality monitoring system in Taiwan from 2012 to 2017. These models showed better performance compared to traditional forecasting approaches. The results were evaluated using statistical measures such as RMSE, MAE, MSE, and R^2, which confirms their higher accuracy and usefulness in air quality forecasting.

In Won et al. (2021), the hygroscopic properties of PM, which affect light scattering and absorption are analyzed. This study explores the behavior of coarse PM (CPM) and fine PM ($PM_{2.5}$) under varying weather conditions and their influence on visibility. A censored regression model was developed to analyze the relationship between PM concentrations and meteorological factors, enabling the calculation of the optical hygroscopic growth factor and hygroscopic mass growth. These metrics were applied to $PM_{2.5}$ field measurements using low-cost sensors in two distinct regions. The results indicate that CPM and $PM_{2.5}$ concentrations negatively impact visibility, with relative humidity (RH) significantly modulating these effects. The modeled hygroscopic growth factors aligned closely with observed values, particularly under haze and mist conditions. Adjusting $PM_{2.5}$ concentrations for RH based on visibility-derived growth factors notably improved the accuracy of low-cost PM sensors. This study underscores the importance of accounting for PM-meteorology interactions in both visibility forecasting and sensor calibration.

The paper presented by Kowalski et al. (2020) evaluates the feasibility of using modern neural networks to forecast PM_{10} concentrations over a 24-hour horizon. This paper analyzes various error measures and compares them with other prediction methods. The results demonstrate that the proposed convolutional neural network can be an effective tool for detailed air quality forecasts.

In the work depicted by Tong et al. (2024), a transformer-based model for spatiotemporal forecasting of $PM_{2.5}$ concentrations is proposed. Using aerosol optical depth data from California, the model outperforms LSTM and other baselines across multiple error metrics, demonstrating strong robustness and stability in complex prediction scenarios.

The study described by Nidzgorska-Lencewicz (2018) used artificial neural networks to forecast hourly PM_{10} concentrations in the Tricity area (Poland) during the winter periods of 2002/2003–2016/2017. Models based on measurement data and meteorological factors achieved satisfactory results (R^2 ranging from 0.452 to 0.848), confirming the effectiveness of ANNs in predicting air pollution levels.

The paper presented by Qin et al. (2014) deals with the forecasting of PM concentrations to support regulatory planning, emission reduction, and early warning system implementation. The authors propose the CS-EEMD-BPANN model, which combines gray correlation analysis (GCA), ensemble empirical mode decomposition (EEMD), cuckoo search algorithm (CS), and backpropagation neural networks (BPANN). The unique element is the use of GCA to identify potential predictors of PM among other air pollutants (CO, NO_2, O_3, SO_2) and meteorological variables (wind speed and direction, temperature, humidity, pressure). The analysis results indicate that CO, NO_2, and SO_2 are strongly related to PM, and including these predictors significantly improves the model performance, which confirms the effectiveness of the proposed approach.

In Czernecki et al. (2021), the application of machine learning methods to short--term forecasting of PM_{10} and $PM_{2.5}$ concentrations in four large Polish agglomerations, based on 10 years of measurement data from background, traffic, and industrial stations is presented. Four models were tested: stepwise regression, random forests, XGBoost, and neural networks, demonstrating that XGBoost performed best, while stepwise regression performed worst. The results confirm the high utility of machine learning algorithms in predicting air pollution episodes and emphasize the importance of meteorological variables in air quality modeling.

The study described by Rogula-Kozłowska et al. (2014) examines seasonal changes in $PM_{2.5}$ concentration and chemical composition at three locations in Poland (Katowice, Gdańsk, and Diabla Góra). The results indicate very high $PM_{2.5}$ concentrations, especially in winter in Katowice, where the dominant components were carbon compounds, secondary aerosols, and benzo(*a*)pyrene (BaP), posing the greatest health risk during the heating season.

In the work presented by Vovk et al. (2024), a hybrid approach that combines the EMEP4PL chemical transport model with random forest to improve the accuracy of $PM_{2.5}$ estimates in Poland, where CTMs alone often underestimate concentrations, is proposed. The model integrates outputs with meteorological, emission, land use, and temporal predictors, and achieved substantially higher accuracy ($R^2 = 0.71$ vs. 0.38 for EMEP4PL), reduced bias, and better detection of severe pollution episodes.

The study described in Tariq et al. (2023) addresses the challenge of forecasting fine particulate matter ($PM_{2.5}$) concentrations in underground subway stations, where air quality is strongly affected by nonlinear dynamics and uncertainty. The authors propose a hybrid framework that combines a gated probabilistic transformer with a genetic algorithm-based quantile scheduling approach, improving both early warning accuracy and ventilation control. Compared to conventional methods, their model achieves narrower prediction intervals, higher accuracy in sequential health risk assessment, and measurable reductions in $PM_{2.5}$ levels, demonstrating its potential for smart building applications.

The study by Kujawska et al. (2022) compares various machine learning models for forecasting PM_{10} concentrations in Lublin, using data from 2017–2019, including meteorological and chemical parameters. The artificial neural network achieved the best results, achieving R = 0.89, effectively predicting the risk of PM_{10} exceedances.

In Cichowicz et al. (2020), the analysis of PM_{10} concentrations in the winter (December–January) in 2009–2015 in Poland showed that low wind speeds are conducive to higher levels of air pollution. The study included data from three monitoring stations in the region, enabling the identification of smog episodes (so-called black smog) and verification of measurements at the station located near a large incineration plant. The results confirm the influence of wind direction on the movement of pollutants or their local deposition.

In the paper presented by Zeng et al. (2023), a hybrid forecasting model that combines CEEMDAN decomposition with a deep transformer architecture to improve long-term $PM_{2.5}$ prediction is proposed. By introducing a specialized embedding layer and a non-autoregressive multi-step decoder, the model reduces error accumulation and significantly outperforms conventional RNN-based approaches on public datasets.

The study presented by Kryza et al. (2019) discusses an air quality forecasts for Poland using the WRF-Chem model, comparing the baseline approach (GENEMIS) with the heating degree day (HDD) method. Both approaches yielded similar results, with HDD better representing concentrations on warmer days, while the baseline method tended to overestimate them.

The work by Ramentol et al. (2023) concerns the prediction of nitrogen dioxide NO_2 concentrations using machine learning methods, using historical pollutant data and meteorological variables. The study focuses on the city of Erfurt (Germany) and includes modeling of temporal dependencies using embedded variables that allow the model to take into account local events such as traffic or specific occasions. Experiments show that the proposed model achieves better forecasting accuracy compared to other models, especially during periods when traditional seasonal patterns change, such as holidays.

In Du et al. (2022), a new hybrid $PM_{2.5}$ concentration forecasting system, which also takes into account the evaluation of health effects and economic losses, is discussed. First, an efficient data analysis method was used to extract the main features of the $PM_{2.5}$ dataset, minimizing the influence of noise. Then, the Harris hawk optimization algorithm was introduced to tune the extreme learning machine (ELM) model, which achieved high prediction accuracy. The health effects and economic costs of pollution were estimated based on the predicted $PM_{2.5}$ values. Experiments were conducted on the real data from Beijing, Tianjin, and Shijiazhuang, and the results showed that the proposed system can support environmental management, emission reduction, and health prevention, and can be applied to various fields such as $PM_{2.5}$ health research.

The work proposed by Kouziokas et al. (2020) introduces a novel support vector machine (SVM) kernel by combining the transformed weight vector of a particle swarm neural network (ANN) optimized by the particle swarm algorithm (PSO) into a Bayesian-optimized SVM kernel for the task of forecasting the concentration of PM_{10} as time series. The proposed model shows higher forecast accuracy compared to the traditional ANN and SVM-optimized models, which is confirmed by the experimental results.

The study presented in Sharma et al. (2020) presents a hybrid deep learning CLSTM model combining a convolutional neural network (CNN) with an LSTM network to predict hourly total suspended particle (TSP) concentrations in Australia. The CLSTM model has been extensively tested and shows better performance than an ensemble of five other machine learning models.

PROPOSED METHODS

In this section, we describe the proposed deep learning architectures for air pollutant concentration prediction. The proposed methods include a convolutional-recurrent autoencoder and a hierarchical autoencoder. All the proposed models are used to predict the PM concentration, including the weather conditions, such as the temperature, wind speed, wind direction, and air pressure. More formally, the $\boldsymbol{x}_t$ is the input data, containing PM historical concentration $\boldsymbol{x}^{conc}$, temperature $\boldsymbol{x}^{temp}$ expressed in °C, wind speed $\boldsymbol{x}^{wind}$ (km/h), wind direction $\boldsymbol{x}^{wdir}$ (degrees), and air pressure $\boldsymbol{x}^{pres}$ (hPa) in a time step t, and is defined as follows:

$$\boldsymbol{x}_t = [\boldsymbol{x}^{conc}, \boldsymbol{x}^{temp}, \boldsymbol{x}^{wind}, \boldsymbol{x}^{wdir}, \boldsymbol{x}^{pres}]$$

The output of the model is a tensor of shape (1,1), representing a single forecasted value of PM_{10} or $PM_{2.5}$ concentration at the next time step. Let us define the sigmoid and hyperbolic tangent functions. The sigmoid function is defined as eq. 1 and the hyperbolic tangent function is defined as eq. 2.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{1}$$

$$\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{2}$$

CONVOLUTIONAL-RECURRENT AUTOENCODER

The convolutional-recurrent autoencoder architecture integrates the advantages of convolutional (for spatial feature extraction) and recurrent (for modeling temporal dependencies) layers (Zheng & Zhang, 2023). The model architecture consists of two main parts: an encoder and a decoder. In the encoder part, the data passes through two convolutional (1D) layers that filter local patterns along the time axis within the features, allowing the capture of local dependencies between different variables (e.g. the effect of wind speed on pollutant concentration). The result of this operation is then processed by the xLSTM unit, which models global temporal dependencies, producing internal representations of reduced dimension.

In a classic autoencoder, the goal of the decoder is to reconstruct the input, and the output is a reconstruction of the same input. In the proposed case, the decoder will be configured to generate predictions for the next values instead of reconstructing the entire input sequence. This is realized by utilizing the linear layers used in the decoder, which allow the model to directly predict the values of future time steps. Thus, in this version, the output is the future values of the time series. The model is trained using the mean squared error (MSE) as the loss function, and the optimization procedure seeks to minimize this loss by reducing the difference between the forecasts and the observed values. This allows the model to learn a representation of the features specific to pollution and weather data, which is crucial for successful predictions. In the prediction process, the network receives historical data as input and generates predictions for subsequent values in the time series. The structure of the considered autoencoder is presented as Fig. 1.

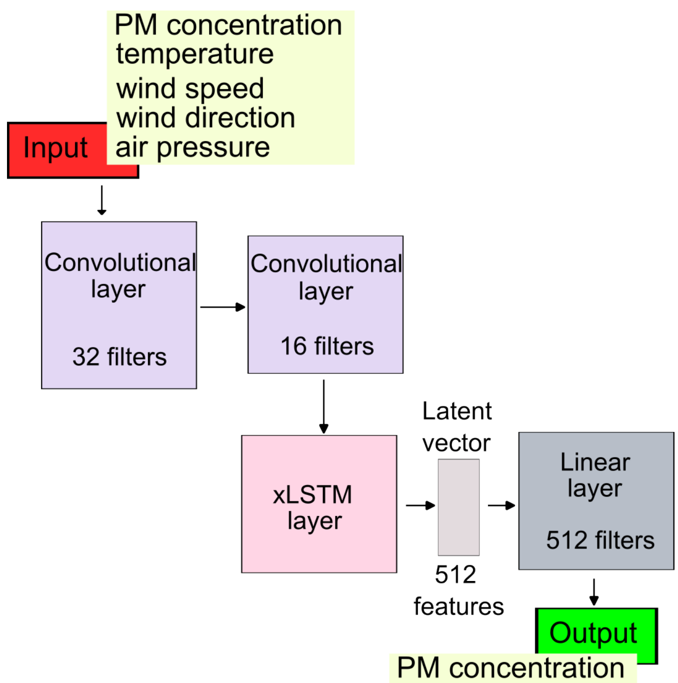


Fig. 1. The structure of the convolutional-recurrent autoencoder (Source: Author's own study)

HIERARCHICAL AUTOENCODER

Hierarchical autoencoder is a deep learning model designed for temporal data analysis and forecasting (Tran et al., 2021). The model uses two main components: an encoder and a decoder, which work together to compress and reconstruct the input data. The encoder transforms the input data, consisting of a time series of air pollutant concentration, temperature, wind speed, and direction values, into a lower-dimensional representation. This process allows for capturing the most relevant features of the data while eliminating redundancy and noise.

The encoder consists of several linear layers interspersed with ReLU activation functions that reduce the dimensions of the input data until a bottleneck is reached. At this point, the data is compressed to its most critical representations. The decoder then reverses the encoding process, transforming the compressed representation back to the original dimensions of the input data using successive linear and activation layers. The final decoder layer applies a sigmoid activation function, which ensures the stability and interpretability of the results.

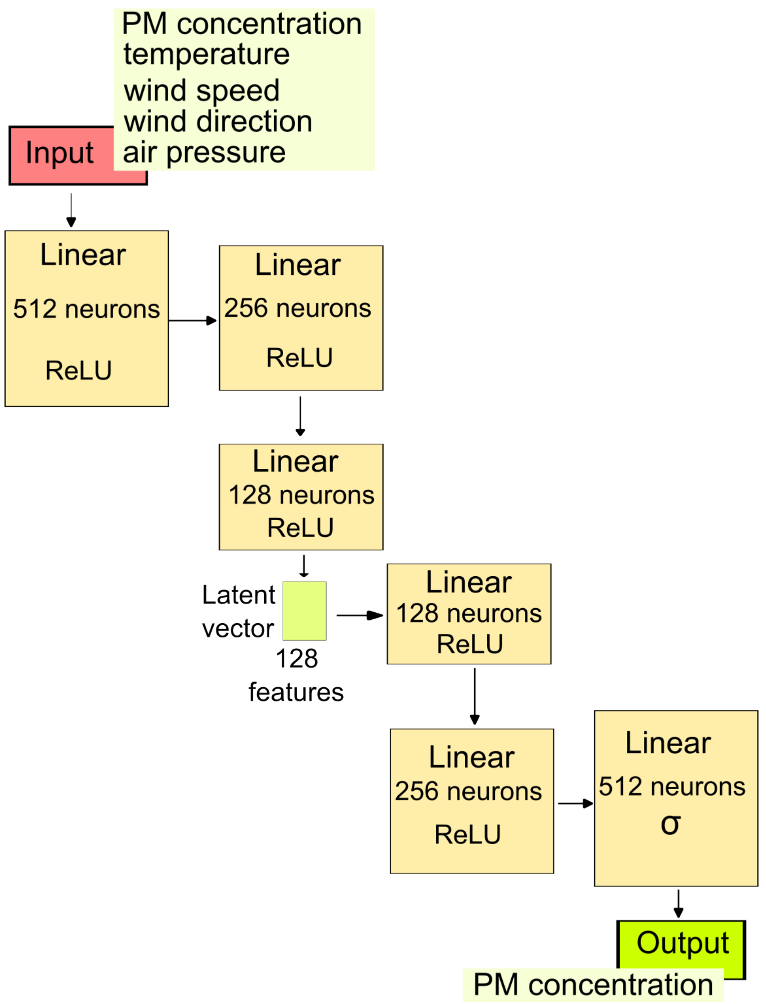


Fig. 2. The structure of the hierarchical autoencoder (Source: Author's own study)

The model is trained using a mean squared error (MSE) loss function, which measures the difference between the original input data and its forecasts. The Adam optimizer updates the network weights, minimizing the forecasting error. The training process involves iteratively tuning the network parameters, allowing the model to efficiently learn a representation of the input data. Once training is complete, the model is able to generate predictions for new data by reconstructing its key features based on the learned structure.

In the case of pollutant concentration forecasting, the model generates output by generating a time sequence that is shifted in time, i.e. the model generates the next value based on the previous one. The graphical structure of the proposed autoencoder is presented as Fig. 2.

EXPERIMENTAL EVALUATION

In this section, we evaluate the performance of the proposed methods through forecasting simulations based on real-world data obtained from the GIOS. This dataset includes historical concentration values (in μg/m^3) for various air pollutants. Specifically, our goal is to train the proposed models using historical $PM_{2.5}$ and PM_{10} concentration data from several Polish cities (referred to as the original dataset), while utilizing relevant meteorological variables. The predicted values are then compared with the recorded concentrations to assess forecasting accuracy. The two proposed models are a convolutional-recurrent autoencoder (AE 1) and a hierarchical autoencoder (AE 2), referred to in parentheses in the evaluation tables for clarity.

Meteorological data were included in the training and consist of the following variables:

- temperature (°C)
- wind speed (km/h)
- wind direction (degrees)
- atmospheric pressure (hPa)

These data were retrieved from the Meteostat platform[2], which aggregates historical weather information from numerous global sources. The meteorological records span the same period as the pollution data – from January 1, 2022, to December 31, 2022. The selected meteorological features were chosen for their completeness and relevance to PM concentration modeling. For example, temperature influences atmospheric chemical reactions; wind affects pollutant dispersion;

[2] https://meteostat.net/en/

and pressure plays a role in temperature inversions that can trap pollutants near the surface, thereby increasing concentration levels. Utilizing these variables allows the models to better reflect the physical processes governing PM behavior. As temperature, we used the daily mean.

For this study, we selected data from the year 2022 because it provided a complete and consistent set of measurements, including both PM concentrations and meteorological variables, necessary for model training and evaluation. Using a single recent year allowed us to thoroughly test and compare the performance of different models under realistic conditions, while avoiding complications from missing or inconsistent data in earlier years. Although 2022 reflects lower pollution levels due to pandemic-related reductions, the methodology is general and can be applied to previous years with higher PM concentrations to assess model robustness. Both the PM concentration data from the GIOS and the meteorological dataset from Meteostat were fully complete, with no missing values, ensuring consistent input for model training and evaluation.

The forecasting models and baseline algorithms were implemented in Python using the PyTorch framework with CUDA support. Statistical analyses were performed using MATLAB. All experiments were conducted on a Gigabyte Aero notebook equipped with an Intel Core i7-13700H CPU, 32 GB of RAM, and an NVIDIA RTX 4070 GPU with 8 GB of VRAM.

COLLECTED DATA AND AREA DESCRIPTION

The evaluation was based on historical concentrations of $PM_{2.5}$ and PM_{10} measured at the air quality monitoring stations in Poland. The locations of the monitoring stations are reported with geographic coordinates (latitude φ, longitude λ) in the WGS84 reference system. The analyzed stations are the following:

- Zielonka Bory Tucholskie (KpZielBoryTu; 53,662117° N; 17,934017° E)
- Zakopane (MpZakopaSien; 49,2936° N; 19,9601° E)
- Częstochowa – SlCzestoZana (50,801918° N; 19,106961° E – $PM_{2.5}$) / SlCzestoBacz (50,836389° N; 19,130111° E – PM_{10})
- Gliwice/Knurów – SlGliwicMewy (50,279481° N; 18,655736° E – $PM_{2.5}$) / SlKnurJedNar (50,233167° N; 18,655722° E – PM_{10})
- Szczecin (ZpSzczec1Maj; 53,712114° N; 16,692517° E)

In Częstochowa and Gliwice/Knurów, $PM_{2.5}$ and PM_{10} are monitored at different stations located in close proximity. For instance, SlCzestoZana and SlCzestoBacz are situated near each other, as are SlGliwicMewy and SlKnurJedNar. While the short distance between these stations suggests a degree of comparability, it should be noted that even small spatial separations (e.g. 500 m) may lead to variations in PM concentrations depending on local emission sources. Figure 3 shows the geographic distribution of the monitoring stations.

The monitoring stations were selected based on the completeness and continuity of their data records for the study period, ensuring no substantial gaps in pollutant or meteorological measurements. In addition, the chosen stations represent diverse environmental contexts in Poland: for instance, Zielonka Bory Tucholskie (KpZielBoryTu) reflects rural background conditions, Zakopane (MpZakopaSien) represents a mountainous area, while Częstochowa (SlCzestoBacz / SlCzestoZana), Gliwice/Knurów (SlGliwicMewy / SlKnurJedNar), and Szczecin (ZpSzczec1Maj) capture urban or industrial settings. This diversity allows the models to be evaluated under heterogeneous air quality conditions and enhances the robustness and generalizability of the proposed forecasting method.

Fig. 3. The location of the air quality monitoring stations in Poland used in the experiment (Source: Author's own study)

TRAINING AND META PARAMETERS

Each model was trained in two configurations. The first uses full input data, including PM concentrations along with temperature, wind speed, wind direction, and air pressure. These inputs are organized into tuples consisting of one value from each variable. Each model takes five consecutive daily observations (i.e. five historical time steps for each variable) as input and generates a single forecast for the next day's PM concentration. The models are trained to minimize the error between the forecasted and observed values using only the PM concentration as the target label.

The second configuration excludes meteorological data and uses only historical PM values as input. Both autoencoders take the five most recent daily observations as input and produce a forecast for the next day. The choice of a five observations was made to effectively capture short-term temporal dependencies in the PM data while maintaining a sufficient number of training samples. Smaller windows would miss important patterns, whereas larger windows would reduce the number of independent samples and increase the risk of overfitting.

Model hyperparameters were selected experimentally. All models were trained using the Adam optimizer with a learning rate of 0.001 and mean squared error (MSE) as the loss function. An early stopping mechanism was applied to prevent overfitting, terminating training once the validation loss no longer decreased; in our experiments, this occurred at around 1000 epochs. The dataset was chronologically divided into three subsets following a 60–20–20 split: the first 60% of the observations were used for training, the next 20% for validation, and the final 20% for independent testing.

All input data (except wind direction) were normalized to the [0, 1] interval. The correctness of normalization was verified by checking the min–max ranges and inspecting histograms before and after scaling. No data distortion was observed (figures omitted for brevity). Specifically, wind direction values (in degrees) were converted into two components, $\sin(\theta)$ and $\cos(\theta)$, where θ is the wind direction angle. This representation avoids the artificial discontinuity between 0° and 359°, which would otherwise be incorrectly mapped to the extreme ends of the interval [0, 1] in standard min–max scaling.

BASE MODEL – STATE-OF-THE-ART ALGORITHM

As a baseline, we used a random forest algorithm (RF), which is an ensemble method based on decision trees. The algorithm creates multiple independent regression trees, each of which learns to predict PM concentrations based on different random subsets of data and features. The final forecast is calculated as the average of all trees' results, reducing the risk of overfitting and improving prediction stability.

The model utilized both historical PM concentration values and meteorological data: temperature, wind speed, wind direction, and atmospheric pressure. Sine and cosine encoding was used for the wind direction variable to preserve its angular nature. The algorithm also uses PM values from previous days/hours as lag features to capture temporal dependencies.

The RF model was run with both historical PM values alone and, in an extended version, with meteorological data (temperature, wind, pressure). This allowed us to compare the impact of atmospheric conditions on the accuracy of particulate matter concentration forecasts.

EVALUATION RESULTS

To evaluate the performance of the forecasts generated by both the proposed and state-of-the-art methods, based on historical PM concentrations and meteorological data, we calculate the following error metrics: mean absolute error (MAE, eq. 3), and mean square error (MSE, eq. 4), which are defined below:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - x_t}{n}\right| \tag{3}$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left(x_t - y_t\right)^2 \tag{4}$$

where $\boldsymbol{x}_t$ is the actual value, $\boldsymbol{y}_t$ is a predicted value, and n is the number of input observations. The results for $PM_{2.5}$ forecasting experiments are presented in Tab. 1.

Tab. 1. Error measure values for the proposed methods from all the air quality monitoring stations for the $PM_{2.5}$ fraction. The AE 1*, AE 2*, and RF* denote the proposed and state-of-the-art methods trained without including the meteorological data (Source: Author's own study)

Air station	Error measure	AE 1	AE 2	AE 1*	AE 2*	RF	RF*
Zielonka Bory Tucholskie	MAE	0.150	0.160	0.230	0.230	0.190	0.210
	MSE	0.040	0.040	0.120	0.110	0.070	0.090
Zakopane	MAE	0.110	0.120	0.190	0.190	0.210	0.230
	MSE	0.020	0.030	0.100	0.100	0.060	0.080
Częstochowa	MAE	0.210	0.190	0.290	0.260	0.220	0.240
	MSE	0.060	0.070	0.140	0.140	0.100	0.120
Gliwice	MAE	0.250	0.210	0.330	0.280	0.240	0.260
	MSE	0.090	0.130	0.170	0.200	0.190	0.210
Szczecin	MAE	0.200	0.220	0.280	0.290	0.250	0.270
	MSE	0.070	0.100	0.150	0.170	0.130	0.150

Analysis of the quality of $PM_{2.5}$ concentration forecasts for five measurement stations shows a clear advantage of autoencoders over random forest. Analysis of the results shows that the autoencoder-based models (AE 1 and AE 2) generally outperform the RF in terms of both MAE and MSE, particularly for the stations in

Zakopane and Zielonka Bory Tucholskie. It can be seen that for most locations, the autoencoders achieve slightly lower forecast errors, suggesting their better ability to capture patterns in the temporal data. When training the models without meteorological data, the error values are higher. The data from the station in Gliwice proved to be the most difficult to model, where all methods generated significantly higher errors, and the RF model, in particular, was characterized by a large increase in MSE. The best results, on the other hand, were obtained in Zakopane, where all methods, especially the autoencoders, showed relatively low MAE and MSE values. Overall, it can be concluded that the proposed AE-based solutions are more stable and outperform the classical random forest algorithm in most cases. Table 2 presents the results of PM_{10} forecasting evaluation.

Tab. 2. Error measure values for the proposed methods from all the air quality monitoring stations for PM_{10} fraction. The AE 1*, AE 2*, and RF* denote the proposed and state-of-the-art methods trained without including the meteorological data (Source: Author's own study)

Air station	Error measure	AE 1	AE 2	AE 1*	AE 2*	RF	RF*
Zielonka Bory Tucholskie	MAE	0.200	0.150	0.280	0.220	0.180	0.200
	MSE	0.050	0.080	0.130	0.150	0.310	0.330
Zakopane	MAE	0.120	0.150	0.200	0.220	0.180	0.200
	MSE	0.020	0.040	0.100	0.110	0.070	0.090
Częstochowa	MAE	0.200	0.180	0.280	0.250	0.210	0.230
	MSE	0.060	0.060	0.140	0.130	0.090	0.110
Knurów	MAE	0.230	0.190	0.310	0.260	0.220	0.240
	MSE	0.070	0.070	0.150	0.160	0.120	0.140
Szczecin	MAE	0.240	0.220	0.320	0.290	0.250	0.270
	MSE	0.080	0.060	0.160	0.130	0.090	0.110

The results clearly show that utilizing meteorological data into the training process improves the accuracy of air quality forecasting. Models AE 1 and AE 2, which used both historical PM concentrations and meteorological variables, achieved lower MAE and MSE error values compared to the versions based solely on historical PM data. Model AE 2 performed particularly well, producing the lowest errors in most locations, especially in Knurów and Szczecin. Importantly, random forest without meteorological data performed worse than the version with additional input variables, with differences particularly visible in Zakopane and Szczecin. The most difficult conditions to predict were in Szczecin, where all models achieved higher errors, although the use of meteorological data continued to reduce MAE and MSE. The best forecasts were observed for Zakopane, where MSE values were the lowest among all stations. In summary, integrating meteorological data increases forecast accuracy and clearly supports autoencoder models, which prove to be more effective than the classic random forest.

CONCLUSION

In this study, two deep learning architectures based on autoencoders were proposed for forecasting PM concentrations: a convolutional-recurrent autoencoder and a hierarchical autoencoder. Both models were evaluated using historical data on $PM_{2.5}$ and PM_{10} concentrations from several monitoring stations. To assess the impact of additional inputs, the models were tested in two configurations: one using only historical PM data, and another using meteorological variables such as temperature, wind, and humidity. The experimental results clearly indicate that the inclusion of weather-related features significantly improves forecasting accuracy across all evaluated metrics and locations. Both proposed models achieved strong performance, with the weather-informed versions consistently outperforming their simplified counterparts. These findings confirm the importance of meteorological context in modeling air pollution dynamics and demonstrate the effectiveness of deep autoencoder-based approaches in this domain.

Future work will focus on extending the proposed models by integrating additional environmental variables such as solar radiation, as well as testing their performance in different geographical regions. Moreover, a comprehensive comparison with other state-of-the-art forecasting methods is planned to further validate the robustness and generalizability of the proposed approach.

REFERENCES

Borah, J., Nadzir, M.S.M., Cayetano, M.G., Majumdar, S., Ghayvat, H., & Srivastava, G. (2024). AiCareAir: Hybrid-ensemble Internet of Things sensing unit model for air pollutant control. *IEEE Sensors Journal*, *99*. https://doi.org/10.1109/JSEN.2024.3397735

Cichowicz, R., Wielgosiński, G., & Fetter, W. (2020). Effect of wind speed on the level of particulate matter PM_{10} concentration in atmospheric air during winter season in vicinity of large combustion plant. *Journal of Atmospheric Chemistry*, *77*, 35–48. https://doi.org/10.1007/s10874-020-09401-w

Czernecki, B., Marosz, M., & Jędruszkiewicz, J. (2021). Assessment of machine learning algorithms in short-term forecasting of PM_{10} and $PM_{2.5}$ concentrations in selected Polish agglomerations. *Aerosol and Air Quality Research*, *21*, 200586. https://doi.org/10.4209/aaqr.200586

Du, P., Wang, J., Yang, W., & Niu, T. (2022). A novel hybrid fine particulate matter ($PM_{2.5}$) forecasting and its further application system: Case studies in China. *Journal of Forecasting*, *41*, 64–85. https://doi.org/10.1002/for.2785

Gryech, I., Asaad, C., Ghogho, M., & Kobbane, A. (2024). Applications of machine learning & Internet of Things for outdoor air pollution monitoring and prediction: A systematic literature review. *Engineering Applications of Artificial Intelligence*, *137*, 109182. https://doi.org/10.1016/j.engappai.2024.109182

Harishkumar, K., Yogesh, K., Gad, I., & Doreswamy, N. (2020). Forecasting air pollution particulate matter ($PM_{2.5}$) using machine learning regression models. *Procedia Computer Science*, *171*, 20572066. https://doi.org/10.1016/j.procs.2020.04.221

Jasiński, R., Galant-Gołębiowska, M., Nowak, M., Ginter, M., Kurzawska, P., Kurtyka, K., & Maciejewska, M. (2021). Case study of pollution with particulate matter in selected locations of Polish cities. *Energies*, *14*(9), 2529. https://doi.org/10.3390/en14092529

Kouziokas, G.N. (2020). SVM kernel based on particle swarm optimized vector and Bayesian optimized SVM in atmospheric particulate matter forecasting. *Applied Soft Computing*, *93*, 106410. https://doi.org/10.1016/j.asoc.2020.106410

Kowalski, P., Sapała, K., & Warchałowski, W. (2020). PM_{10} forecasting through applying convolution neural network techniques. *International Journal of Environmental Impacts*, *3*(1), 31–43. https://doi.org/10.2495/EI-V3-N1-31-43

Kryza, M., Werner, M., & Dore, A.-J. (2019). Application of degree-day factors for residential emission estimate and air quality forecasting. *International Journal of Environment and Pollution*, *65*(4), 325–336. https://doi.org/10.1504/IJEP.2019.103748

Kujawska, J., Kulisz, M., Oleszczuk, P., & Cel, W. (2022). Machine learning methods to forecast the concentration of PM_{10} in Lublin, Poland. *Energies*, *15*(17), 6428. https://doi.org/10.3390/en15176428

Li, T., Hua, M., & Wu, X. (2020). A hybrid CNN-LSTM model for forecasting particulate matter (PM2.5). *IEEE Access*, *8*, 2693326940. https://doi.org/10.1109/ACCESS.2020.2971348

Mauricio-Alvarez, L.-E., Aceves-Fernandez, M.-A., Pedraza-Ortega, J.-C., & Ramos-Arreguin, J.-M. (2024). Evaluation of a transformer-based model for the temporal forecast of coarse particulate matter (PMCO) concentrations. *Earth Science Informatics*, *17*, 3095–3110. https://doi.org/10.1007/s12145-024-01330-6

Nidzgorska-Lencewicz, J. (2018). Application of artificial neural networks in the prediction of PM_{10} levels in the winter months: A case study in the Tricity agglomeration, Poland. *Atmosphere*, *9*(6), 203. https://doi.org/10.3390/atmos9060203

Penkała, M., Rogula-Kozłowska, W., Ogrodnik, P., Bihałowicz, J.S., & Iwanicka, N. (2023). Exploring the relationship between particulate matter emission and the construction material of road surface: Case study of highways and motorways in Poland. *Materials*, *16*, 1200. https://doi.org/10.3390/ma16031200

Polichetti, G., Cocco, S., Spinali, A., Trimarco, V., & Nunziata, A. (2009). Effects of particulate matter (PM_{10}, $PM_{2.5}$ and PM_1) on the cardiovascular system. *Toxicology*, *261*, 1–8. https://doi.org/10.1016/j.tox.2009.04.035

Połednik, B. (2022). Emissions of air pollution in industrial and rural region in Poland and health impacts. *Journal of Ecological Engineering*, *23*, 250258. https://doi.org/10.12911/22998993/151986

Ramentol, E., Grimm, S., Stinzendorfer, M., & Wagner, A. (2023). Short-term air pollution forecasting using embeddings in neural networks. *Atmosphere*, *14*, 298. https://doi.org/10.3390/atmos14020298

Rogula-Kozłowska, W., Klejnowski, K., Rogula-Kopiec, P., Ośródka, L., Krajny, E., Błaszczak, B., & Mathews, B. (2014). Spatial and seasonal variability of the mass concentration and chemical composition of $PM_{2.5}$ in Poland. *Air Quality, Atmosphere & Earth*, *7*, 41–58. https://doi.org/10.1007/s11869-013-0222-y

Sharma, E., Deo, R.C., Prasad, R., Parisi, A.V., & Raj, N. (2020). Deep air quality forecasts: Suspended particulate matter modeling with convolutional neural and long short-term memory networks. *IEEE Access*, *8*, 209503–209516. https://doi.org/10.1109/ACCESS.2020.3039002

Sowka, I., Chlebowska-Styś, A., Pachurka, Ł., Rogula-Kozłowska, W., & Mathews, B. (2019). Analysis of particulate matter concentration variability and origin in selected urban areas in Poland. *Sustainability*, *11*, 5735. https://doi.org/10.3390/su11205735

Swetha, G., Datla, R., Vishnu, C., & Mohan, C.K. (2024). M2-APNet: A multimodal deep learning network to predict major air pollutants from temporal satellite images. *Journal of Applied Remote Sensing*, *18*, 012005–012005. https://doi.org/10.1117/1.JRS.18.012005

Tariq, S., Loy-Benitez, J., & Yoo, C. (2023). Enhancing the sustainable management of fine particulate matter-related health risks at subway stations through sequential forecast and gated probabilistic transformer. *Building and Environment*, *244*, 110780. https://doi.org/10.1016/j.buildenv.2023.110780

Tong, W., Limperis, J., Hamza-Lup, F., Hu, Y., & Li, L. (2024). Robust transformer-based model for spatiotemporal $PM_{2.5}$ prediction in California. *Earth Science Informatics*, *17*, 315–328. https://doi.org/10.1007/s12145-023-01138-w

Tran, D., Nguyen, H., Tran, B., La Vecchia, C., Luu, H.N., & Nguyen, T. (2021). Fast and precise single-cell data analysis using a hierarchical autoencoder. *Nature Communications*, *12*, 1029. https://doi.org/10.1038/s41467-021-21312-2

Qin, S., Liu, F., Wang, J., & Sun, B. (2014). Analysis and forecasting of the particulate matter (PM) concentration levels over four major cities of China using hybrid models. *Atmospheric Environment*, *98*, 665–675. https://doi.org/10.1016/j.atmosenv.2014.09.046

Vovk, T., Kryza, M., & Werner, M. (2024). Using random forest to improve EMEP4PL model estimates of daily $PM_{2.5}$ in Poland. *Atmospheric Environment*, *332*, 120615. https://doi.org/10.1016/j.atmosenv.2024.120615

Won, W.-S., Oh, R., Lee, W., Ku, S., Su, P.-C., & Yoon, Y.-J. (2021). Hygroscopic properties of particulate matter and effects of their interactions with weather on visibility. *Scientific Reports*, *11*, 16401. https://doi.org/10.1038/s41598-021-95834-6

Zeng, Q., Wang, L., Zhu, S., Gao, Y., Qiu, X., & Chen, L. (2023). Long-term $PM_{2.5}$ concentrations forecasting using CEEMDAN and deep transformer neural network. *Atmospheric Pollution Research*, *14*(9), 101839. https://doi.org/10.1016/j.apr.2023.101839

Zheng, Z., & Zhang, Z. (2023). A temporal convolutional recurrent autoencoder based framework for compressing time series data. *Applied Soft Computing*, *147*, 110797. https://doi.org/10.1016/j.asoc.2023.110797

https://meteostat.net/en/

https://www.gov.pl/web/gios-en

PUBLICATION INFO		
SUBMITTED: 2025.07.23	ACCEPTED: 2025.10.20	PUBLISHED ONLINE: 2025.12.5

A.6 Publikacja konferencyjna A6 – Załącznik 6

Forecasting Particulate Matter with Neural Networks: Accuracy and Explainability Using Meteorological Data

Jarosław Bernacki[1][0000-0002-4488-3488],
Izabela Sówka[2][0000-0001-9810-4673]

[1] Faculty of Information and Communication Technology
[2] Faculty of Environmental Engineering
Wrocław University of Science and Technology
Wyb. Wyspiańskiego 27, Wrocław, Poland
firstname.lastname@pwr.edu.pl

Abstract. In recent years, forecasting particulate matter (PM) concentrations has become a key tool in air quality monitoring and environmental decision-making. This paper compares modern neural network methods for forecasting PM concentrations using both meteorological data and PM history alone. The models were assessed for forecast accuracy using metrics such as MAE, RMSE, and R^2, and their interpretability was analyzed using LIME and SHAP methods to assess the impact of individual meteorological variables on forecast results.

Keywords: forecasting · predicting · deep models · particulate matter

1 Introduction

Forecasting particulate matter (PM) concentrations is essential in public health protection and air quality management, particularly in urban areas where pollution is high. PM fractions, including PM_1, $PM_{2.5}$, and PM_{10}, are associated with respiratory and cardiovascular diseases, and long-term exposure negatively impacts quality of life [3, 14]. Accurate PM forecasts allow both to warn about high pollution levels and also support environmental policy planning and emission reduction measures. Various forecasting methods are used in the literature, ranging from traditional statistical and regression models to modern machine and deep learning algorithms that can capture nonlinear and multivariate relationships in temporal and meteorological data [16, 22, 23].

In this paper, we focus on evaluating three modern neural network models for forecasting PM concentrations: temporal fusion transformer (TFT), Kolmogorov–Arnold Network (KAN), and temporal convolutional network (TCN). The proposed methods are validated using data from multiple air quality monitoring stations located in different regions of Poland, representing diverse environmental and pollution conditions, ranging from coastal and low-emission areas to highly polluted urban and mountainous regions. Each model is trained and tested in two scenarios: using only historical PM concentrations as input, and using both PM concentrations and meteorological variables (wind speed, wind direction, and temperature) to evaluate the added value of environmental data. TFT combines recurrent layers with a self-attention mechanism to capture long- and short-term dependencies and enables variable selection. KAN utilizes the Kolmogorov–Arnold representation theorem to approximate complex nonlinear relationships through fully connected layers with heterogeneous nonlinear activations. TCN employs dilated causal convolutions with residual connections to efficiently model long-range temporal dependencies. For comparison, two baseline models are also evaluated: a standard RNN and an LSTM network. We compare all models for forecast accuracy using metrics such as MAE, RMSE, and R^2, and analyze their interpretability using LIME and SHAP methods to assess the impact of individual meteorological variables on the prediction results. This approach allows both to determine the efficacy of different neural models and to understand the extent to which meteorological data contribute to improving forecast quality, which can support decision-making in air quality monitoring and management.

1.1 Contribution

The contribution of this paper is twofold. First, we propose and evaluate modern neural network methods for forecasting PM concentrations, including temporal fusion transformer, Kolmogorov–Arnold Network, and temporal convolutional network, considering both univariate and multivariate inputs with meteorological variables. Second, we analyze the impact of individual meteorological features on the predictions using interpretability methods such as LIME and SHAP, providing insights into which environmental factors most significantly influence forecast accuracy.

1.2 Organization of the paper

The paper is organized as follows. The next section describes previous and related works. Section 3 depicts the neural networks for PM forecasting. Section 4 discusses the preliminaries for the methods' evaluation. In section 5, the experimental results of the proposed methods are described. Finally, the last section concludes this work.

2 Previous and related work

Particulate matter concentrations are predicted using a variety of methods, including traditional statistical models and modern approaches based on machine learning and deep learning. Regression [1] and temporal models capture seasonal and meteorological variations [7], while algorithms such as random forest [5], SVM [10], LSTM [21], and hybrid TCN-LSTM combinations [19] enable accurate hourly or multiday forecasting. Many studies also utilize variable sensitivity analysis to improve modeling efficiency and identify the most important factors influencing PM concentrations. These diverse approaches support more effective air quality monitoring and environmental decision-making.

The study [8] shows that PM_{10} concentration in may depend on meteorological conditions and contributes to atmospheric warming. Multiple linear regression (MLR), random forest (RF), XGBoost, and artificial neural networks (ANN) models were successfully used to forecast it. Including PM_{10} concentration from the previous day significantly improves forecast accuracy (by 57–92%), with MLR best describing seasonal variability, RF best predicting the following day, and ANN the next 2–3 days. In [4], the relationships between vehicle emissions, meteorological conditions, and $PM_{2.5}$ concentrations are analyzed. In the study [9], highly accurate PM_{10} and $PM_{2.5}$ concentration forecasts were obtained using meteorological data and machine learning algorithms based on decision trees.

The research [15] evaluates the performance of three machine learning algorithms for forecasting $PM_{2.5}$ and PM_{10} concentrations in metropolitan areas: SVM, random forest, and k-NN, paying particular attention to the impact of the training set on model accuracy. Analyses show that random forest often achieves higher R^2 than SVM, k-NN is effective in data preprocessing, and hybrid models outperform approaches based on single algorithms. In [17], a hybrid TCN-LSTM model to forecast PM concentrations is proposed. Sensitivity analysis identified key factors influencing PM (O_3 for $PM_{2.5}$, $PM_{2.5}$ for PM_{10}, CO for both) and showed that eliminating irrelevant variables increases modeling efficiency and improves air quality monitoring. The study [6] presents deep learning models (LSTM, RNN, GRU) for forecasting hourly $PM_{2.5}$ concentrations, using air quality data and meteorological factors. The LSTM+LSTM model achieved the best result in terms of R^2 measure and showed good generalization ability through different stations in the city, outperforming other studies in the literature. In [18], a hybrid approach combining input variable selection, machine learning, and regression was proposed to forecast PM_1, $PM_{2.5}$, PM_{10}, and AQI concentrations, achieving high accuracy ($R^2 > 0.95$). Additionally, 13 multivariate models and a multistep forecasting application using NARMAX and decision trees were developed, achieving R^2 exceeding 0.93 and confirming high forecast precision.

3 Neural network architectures for PM forecasting

In this chapter, we describe the proposed approach for forecasting air pollutant concentrations using time series data. We first present the formal foundations of h-step ahead forecasting, considering both historical PM measurements and their extension to include meteorological variables such as wind speed, wind direction, and temperature. Building on this framework, we introduce three modern neural network models employed in this study: temporal fusion transformer, Kolmogorov–Arnold Network, and temporal convolutional network. For each model, we describe the architecture and mechanisms used to capture temporal dependencies and improve predictive performance.

3.1 Problem formulation and forecasting task definition

A time series is defined as an ordered sequence of observations of a certain quantity at subsequent time points. In this paper, a time series will be understood as all observations of concentrations of selected air pollutants and their associated meteorological parameters, collected at subsequent time points at individual measurement stations. Formally, the time series $\mathbf{y}$ can be defined as follows. Let:

- $\{\mathbf{y}\}_{t=1}^{T}$ be the observed time series (containing historical air pollutant concentration values at successive time intervals);
- t – time index (e.g., hour, day);
- T – last known observation time.

In practice, a time series can therefore be a vector of the following form:

$$\mathbf{y} = [y_1, y_2, \ldots, y_T], \tag{1}$$

where y_T denotes the observation value in time unit t. In addition to the series $\mathbf{y}$ describing the concentration of a given pollutant (e.g., $\mathrm{PM}_{2.5}$), a vector of meteorological variables is also considered:

$$\mathbf{x}_t = [v_t, d_t, T_t], \tag{2}$$

where v_t is the wind speed, d_t is the wind direction, and T_t is the air temperature at the moment t. These variables create a multidimensional time series:

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T]. \tag{3}$$

The h-step ahead forecasting task involves estimating future pollutant concentration values, taking into account both historical pollutant data and meteorological variables:

$$\hat{y}_{T+h|T} = f(\mathbf{y}_{1:T}, \mathbf{X}_{1:T}), \quad \text{for} \quad h = 1, 2, \ldots, H, \tag{4}$$

where:

- $\hat{y}_{T+h|T}$ – forecast of the value of y_{T+h} based on information available up to and including time T;
- H – forecasting horizon (number of steps forward to be forecasted);
- $f(\cdot)$ – nonlinear function approximated by the applied forecasting model.

3.2 Temporal fusion transformer

The temporal fusion transformer (TFT) model was designed as a hybrid architecture combining recurrent neural networks with an attention mechanism, enabling simultaneous modeling of short-term and long-term temporal dependencies and evaluation of the importance of input variables.

Input and embedding layer Each input variable (PM concentrations and meteorological variables) is mapped to a common feature space using embedding and normalization layers. This allows for unifying the data scale and efficiently combining information from different sources.

Variable selection network A variable selection network based on gated residual networks (GRN) is used for each time instant, assigning weights to individual input features. This module allows for dynamically determining which variables (PM or meteorological) have the greatest impact on the forecast at a given time step.

LSTM block We propose to use two LSTM layers in the encoder and one LSTM layer in the decoder, each with a hidden dimension of 128 neurons. The encoder processes a sequence of historical data, learning short- and medium-term dependencies. The decoder generates representations for the forecast horizon, ensuring temporal continuity between past observations and future predictions.

Temporal self-attention A multi-head attention mechanism is implemented at the output of the LSTM block, which analyzes the relationships between all time steps in the sequence. This mechanism allows the model to: identify key moments in the past (e.g., high PM episodes), model long-term relationships, and differentiate the influence of observations over time using adaptive attention weights. Formally, for each attention head, the query (Q), key (K), and value (V) matrices are computed:

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5}$$

which allows for the determination of weights defining the importance of individual time steps for the current forecast.

Output layer The outputs from the attention mechanism are passed to the fully connected layer (dense), which generates the final PM concentration forecast for one or more time steps ahead. The model supports both single-step $(t+1)$ and multi-step $(t+1, \cdots, t+H)$ forecasting.

3.3 Kolmogorov-Arnold Network

Kolmogorov–Arnold Networks (KAN), inspired by the Kolmogorov–Arnold representation theorem, were first introduced in [11]. The theorem states that any continuous multivariate function can be represented as a superposition of continuous univariate functions, which forms the theoretical foundation of the KAN architecture. Unlike conventional multilayer perceptrons, KAN explicitly models nonlinear transformations through learnable univariate functions, enabling more flexible approximation of complex dependencies.

In this study, a modified KAN architecture is employed to better adapt the model to time series forecasting of PM concentrations. The network is composed of fully connected layers with heterogeneous nonlinear basis functions and regularization mechanisms. The proposed KAN structure is defined as follows:

(1) A first fully connected layer (FC1) with 128 neurons;
(2) A second fully connected layer (FC2) with 256 neurons;
(3) A third fully connected layer (FC3) with 128 neurons;
(4) A fourth fully connected layer (FC4) with 64 neurons;
(5) A final output layer producing the PM concentration prediction.

Unlike traditional neural network architectures that utilize predefined activation functions, KAN models nonlinear relationships through trainable one-dimensional functions associated with network connections. In our implementation, these functions are represented as linear combinations of B-spline basis functions, evaluated using the Cox–de Boor recursive algorithm [12].

The network processes the input time series using a sliding window approach. Each window consists of historical PM values and, when applicable, meteorological variables. The extracted windows are flattened and passed through the network layers, enabling the model to learn nonlinear temporal dependencies without explicit recurrent or convolutional operations. To prevent overfitting, FC1–FC4 layers are followed by dropout. The window size is treated as a tunable hyperparameter and optimized during model training.

3.4 Temporal convolutional network

Temporal convolutional networks (TCN) are convolutional architectures specifically designed for sequential data processing, employing dilated causal convolutions to model long-range temporal dependencies while avoiding the gradient instability issues commonly observed in recurrent neural networks [2]. The receptive field of a TCN grows exponentially with network depth through the use of increasing dilation factors, enabling efficient learning of both short- and long-term temporal patterns.

In this study, an enhanced TCN architecture based on residual blocks is employed to improve representational capacity and training stability. Each residual block consists of two one-dimensional convolutional layers with identical dilation rates, followed by nonlinear activation and dropout regularization. Skip connections are introduced to facilitate gradient flow and to preserve low-level temporal features. The proposed TCN architecture is defined as follows:

(1) Five residual TCN blocks, each composed of two 1D convolutional layers with kernel size 3, SELU activation, layer normalization, and dropout. The dilation rates of successive blocks are set to $\{1, 2, 4, 8, 16\}$;
(2) A temporal attention layer applied to weight the importance of time steps within the receptive field;
(3) A final 1×1 convolutional layer that maps the extracted features to the target PM concentration.

The dilation schedule $1, 2, 4, 8, 16$ ensures that the receptive field covers the entire input window, allowing the model to learn multiscale temporal dependencies. The attention mechanism further enhances the model by adaptively focusing on the most informative historical time steps, particularly during pollution episodes. The sliding window is used to generate input sequences, enabling multi-step forecasting based on historical PM and meteorological variables.

4 Experimental evaluation – preliminaries

In this section, we theoretically introduce the experimental evaluation. We discuss the utilized data, implementation details, training protocol, implementation details, used baseline models, as well as error measures and model explainability methods, including LIME and SHAP techniques.

4.1 Collected data and area description

We evaluate the performance of the proposed methods through forecasting simulations based on real-world data obtained from the Chief Inspectorate of Environmental Protection (GIOS)[3]. This dataset includes historical concentration values (in $\mu g/m^3$) for various air pollutants. Specifically, we train and test the proposed forecasting models: temporal fusion transformer (TFT), Kolmogorov–Arnold Network (KAN), and temporal convolutional network (TCN). The forecasting evaluation was conducted with the $PM_{2.5}$ and PM_{10} historical concentrations. They were analyzed in the following air quality monitoring stations in Poland: Bory Tucholskie (Zielonka), Zakopane, Częstochowa, Gliwice, and Szczecin. The data cover daily particulate matter concentration measurements from January 1, 2022, to December 31, 2022. The models are trained in a twofold manner:

(1) Using only historical $PM_{2.5}$ and PM_{10} concentration data;
(2) Using historical PM concentrations augmented with meteorological variables.

This allows to directly compare forecasting accuracy between models trained with and without meteorological information and to quantify the impact of weather-related factors on prediction performance. The predicted values are then compared with the recorded concentrations to assess forecasting accuracy. Meteorological data included in the training process consist of the following variables:

- Temperature (°C);
- Wind speed (km/h);
- Wind direction (degrees).

The meteorological data were obtained from the Meteostat platform[4], which collects and harmonizes historical weather observations from a wide range of international sources. The weather records correspond to the same time interval

[3] https://www.gov.pl/web/gios-en
[4] https://meteostat.net/en/

as the PM measurements, covering the daily period from January 1, 2022, to December 31, 2022. The meteorological variables were selected based on their availability and their known influence on PM dynamics. In particular, air temperature affects atmospheric chemical processes, wind conditions determine the transport and dilution of pollutants, and atmospheric pressure is associated with thermal inversions that may limit vertical air mixing and lead to pollutant accumulation near the ground. Including these variables enables the forecasting models to more realistically represent the physical mechanisms governing PM concentrations. Daily mean temperature values were used in the experiments.

4.2 Implementation

The proposed models and comparative methods were implemented in Python using the PyTorch framework with NVIDIA CUDA acceleration. Statistical analyses were conducted in MATLAB. All experiments were performed on a workstation equipped with an Intel Core i5-10400 processor, 32 GB of RAM, and an NVIDIA RTX 3060 GPU with 12 GB of video memory.

4.3 Training protocol

Each model was trained under two different setups. In the first setup, only historical PM values were used as input. For TFT, KAN, and TCN, the five most recent daily PM observations served as input to predict the next day's concentration. The choice of a five-day window balances capturing short-term temporal dependencies with maintaining a sufficient number of independent training samples: shorter windows risk missing relevant patterns, while longer windows reduce the number of effective samples and may increase overfitting. The second setup uses full input data, combining PM concentrations with meteorological variables, i.e., temperature, wind speed, and wind direction. These inputs are represented as tuples containing one value per variable. The models process five consecutive daily observations for each variable to produce a forecast for the following day. Only the PM values are used as targets, allowing the models to minimize the prediction error effectively.

Model hyperparameters were determined experimentally. All models were trained with the Adam optimizer (learning rate 0.001) using mean squared error (MSE) as the loss function. Early stopping was applied to prevent overfitting, halting training when the validation loss ceased improving, which in our experiments occurred around 500 epochs. The dataset was split chronologically into training, validation, and testing sets following a 60–20–20 scheme: the first 60% for training, the next 20% for validation, and the remaining 20% for independent testing.

All input variables, except wind direction, were normalized to the $[0, 1]$ range. The correctness of normalization was verified by inspecting min–max ranges and histograms before and after scaling. Wind direction, measured in degrees, was encoded as two components, $\sin(\theta)$ and $\cos(\theta)$, where θ represents the wind angle. This approach avoids the discontinuity between 0° and 359°, which would otherwise be misrepresented by standard min–max scaling.

4.4 Baseline models

We use two baseline models: a recurrent neural network (RNN) and long short-term memory (LSTM), which, due to paper limitations, we briefly recall below.

RNN A standard recurrent neural network (RNN) with a single hidden layer of 50 neurons is used to model temporal dependencies in historical PM data. It processes sequences step by step, capturing short-term patterns but may struggle with long-term dependencies due to vanishing gradients.

LSTM A long short-term memory (LSTM) network with one hidden layer of 50 units is employed to better capture long-range temporal dependencies. The LSTM's gating mechanisms allow it to retain relevant information over longer sequences compared to an RNN.

4.5 Error measures

To evaluate the performance of the forecasts generated by both the proposed and state-of-the-art methods, we calculate several error metrics: mean absolute error (MAE, eq. 6), root mean square error (RMSE, eq. 6), and R^2 (coefficient of determination, eq. 7).

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - x_t|, \quad \mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - x_t)^2} \tag{6}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_t - \bar{y})^2}{\sum_{i=1}^{n}(x_t - \bar{y})^2} \tag{7}$$

where x_t is the actual value, y_t is a predicted value, and $\bar{y}$ is a mean of the data y_t. The coefficient of determination reaches values $[0, 1]$, where 1 denotes a perfect fit, and 0 denotes no fit.

4.6 Model explainability methods

LIME LIME (Local Interpretable Model-agnostic Explanations) is a model-agnostic technique that explains individual predictions by approximating the behavior of a complex model in the vicinity of a given instance using a simple and interpretable surrogate model, typically linear. The resulting local model coefficients indicate the relative importance of input features for the explained prediction [20].

SHAP SHAP (Shapley Additive Explanations) is an explainable artificial intelligence method based on cooperative game theory. It attributes the prediction of a model to individual input features by computing Shapley values, which quantify the marginal contribution of each feature to the final output. This approach provides both local explanations for single predictions and global insight into feature importance through the dataset [13].

5 Experimental results

We present the evaluation of the proposed models: temporal fusion transformer (TFT), Kolmogorov-Arnold Network (KAN), temporal convolutional network (TCN) against the baseline methods: recurrent neural network (RNN) and long short-term memory network (LSTM).

5.1 Forecasting results

The forecasting results for $PM_{2.5}$ and PM_{10} are presented in Tab. 1, and 2.

Table 1. Forecasting performance for $PM_{2.5}$ concentrations across different stations. TFT, KAN, and TCN are the proposed methods, while RNN and LSTM serve as baseline models. Methods denoted without * were trained using only historical PM data, whereas methods with * were trained with both PM and meteorological variables.

		TFT	TFT*	KAN	KAN*	TCN	TCN*	RNN	RNN*	LSTM	LSTM*
	MAE	0.10	0.09	0.11	0.09	0.10	0.08	0.21	0.19	0.19	0.18
KpZielBoryTu	RMSE	0.25	0.22	0.27	0.21	0.26	0.22	0.28	0.26	0.27	0.25
	R^2	0.88	0.91	0.86	0.91	0.86	0.92	0.84	0.86	0.85	0.87
	MAE	0.12	0.09	0.13	0.10	0.10	0.09	0.22	0.19	0.19	0.18
MpZakopaSien	RMSE	0.25	0.19	0.25	0.21	0.24	0.23	0.27	0.24	0.26	0.24
	R^2	0.87	0.91	0.87	0.91	0.88	0.93	0.85	0.86	0.85	0.86
	MAE	0.12	0.10	0.12	0.11	0.14	0.10	0.19	0.18	0.15	0.14
SlCzestoZana	RMSE	0.23	0.21	0.22	0.19	0.25	0.20	0.28	0.27	0.26	0.24
	R^2	0.84	0.89	0.85	0.91	0.87	0.91	0.84	0.85	0.84	0.86
	MAE	0.13	0.10	0.11	0.08	0.11	0.08	0.20	0.18	0.16	0.14
SlGliwicMewy	RMSE	0.27	0.23	0.25	0.21	0.26	0.22	0.29	0.27	0.25	0.24
	R^2	0.89	0.91	0.87	0.91	0.87	0.92	0.85	0.89	0.84	0.85
	MAE	0.13	0.08	0.14	0.09	0.12	0.11	0.21	0.18	0.17	0.15
ZpSzczec1Maj	RMSE	0.24	0.21	0.27	0.22	0.26	0.22	0.29	0.28	0.28	0.27
	R^2	0.87	0.91	0.85	0.90	0.86	0.92	0.81	0.85	0.82	0.83

The results show that all the proposed neural models consistently outperform the baseline RNN and LSTM models in forecasting $PM_{2.5}$ concentrations across all stations. Training the models in the twofold setup clearly demonstrates the added value of meteorological variables. For all three proposed models, utilizing temperature, wind speed, and wind direction (TFT*, KAN*, TCN*) reduces

MAE and RMSE while increasing R^2 compared to the PM-only setup, highlighting improved predictive accuracy. Among the proposed methods, TFT and TCN generally achieve the lowest MAE and RMSE. The R^2 values for TFT*, KAN*, and TCN* range from 0.91 to 0.93, indicating that these models explain more than 90% of the variance in $PM_{2.5}$ concentrations across the evaluated stations when meteorological data are included. Baseline models (RNN and LSTM) also benefit from adding meteorological variables, but their performance remains inferior to the proposed architectures, with higher errors and lower R^2. These results confirm that the TFT, KAN, and TCN architectures are highly effective for $PM_{2.5}$ forecasting, and that integrating relevant meteorological features further enhances model performance.

Table 2. Forecasting performance for PM_{10} concentrations across different stations. TFT, KAN, and TCN are the proposed methods, while RNN and LSTM serve as baseline models. Methods denoted without * were trained using only historical PM data, whereas methods with * were trained with both PM and meteorological variables.

		TFT	TFT*	KAN	KAN*	TCN	TCN*	RNN	RNN*	LSTM	LSTM*
	MAE	0.11	0.10	0.14	0.09	0.12	0.10	0.22	0.21	0.18	0.16
KpZielBoryTu	RMSE	0.25	0.22	0.26	0.22	0.26	0.21	0.29	0.28	0.26	0.25
	R^2	0.88	0.91	0.88	0.91	0.87	0.90	0.84	0.85	0.85	0.86
	MAE	0.12	0.10	0.13	0.10	0.14	0.09	0.22	0.20	0.23	0.21
MpZakopaSien	RMSE	0.26	0.21	0.25	0.21	0.25	0.22	0.26	0.25	0.25	0.24
	R^2	0.87	0.91	0.85	0.91	0.87	0.92	0.83	0.85	0.84	0.86
	MAE	0.11	0.09	0.10	0.10	0.10	0.09	0.18	0.16	0.16	0.14
SlCzestoBacz	RMSE	0.26	0.21	0.24	0.21	0.26	0.23	0.27	0.25	0.26	0.24
	R^2	0.89	0.89	0.84	0.89	0.85	0.91	0.82	0.83	0.84	0.85
	MAE	0.12	0.09	0.15	0.09	0.13	0.08	0.18	0.16	0.15	0.14
SlKnurJedNar	RMSE	0.26	0.22	0.29	0.22	0.24	0.21	0.29	0.25	0.26	0.22
	R^2	0.88	0.90	0.87	0.91	0.87	0.92	0.82	0.84	0.85	0.87
	MAE	0.13	0.11	0.11	0.09	0.14	0.11	0.17	0.15	0.16	0.13
ZpSzczec1Maj	RMSE	0.25	0.21	0.24	0.22	0.26	0.23	0.28	0.26	0.24	0.23
	R^2	0.87	0.89	0.85	0.91	0.84	0.90	0.81	0.85	0.84	0.86

The results in Table 2 indicate that the proposed neural models (TFT, KAN, and TCN) outperform the baseline RNN and LSTM models in forecasting PM_{10} concentrations on all evaluated stations. Similarly, as with $PM_{2.5}$, training the models in a twofold setup demonstrates that using temperature, wind speed, and wind direction slightly improves predictive performance. In nearly all cases, the MAE and RMSE decrease and R^2 increases when meteorological data are included (TFT*, KAN*, TCN*), highlighting the added value of meteorological inputs. Among the proposed methods, TCN* and KAN* consistently achieve the lowest errors, while R^2 values for TFT*, KAN*, and TCN* typically exceed 0.90, meaning these models explain more than 90% of the variance in PM_{10} concentrations. The baseline models also benefit from adding meteorological features, but their overall accuracy remains lower than of the proposed architectures. These results confirm that TFT, KAN, and TCN are highly effective for PM_{10} forecasting, and that the inclusion of relevant meteorological variables enhances the quality of predictions.

5.2 Model explainability methods

We have evaluated the results of model explainability analyses obtained using LIME and SHAP methods, providing insight into how individual input variables influence the forecasting of PM concentrations.

LIME The results of LIME analysis are presented in Tab. 3.

Table 3. The influence of features on prediction by the LIME method for different deep models.

Feature	TFT	KAN	TCN	RNN	LSTM
PM	−0.216	−0.217	−0.137	−0.136	−0.139
Wind speed	0.005	0.002	0.024	−0.022	−0.021
Wind direction	0.004	0.002	−0.010	0.015	0.017
Temperature	0.004	0.001	0.009	0.013	0.011

Analysis of the model's local explanations using the LIME method showed that the PM feature had a dominant, negative effect on the predictions in all the deep models examined, with weight values ranging from -0.216 for TFT to -0.137 for TCN. These results suggest that higher PM values in the sample lower the predicted PM concentration, which may reflect local variations in the data, such as autocorrelation. Meteorological features such as wind speed, wind direction, and temperature showed a much smaller effect, with weight values ranging from -0.022 to 0.024, with their direction (positive or negative) varying between models. For example, wind speed had a positive effect in TCN (0.024) but a negative effect in RNN and LSTM (-0.022 and -0.021). The TFT and KAN models showed similar feature influence profiles, with minimal contribution from meteorological features, which may indicate limited use of these variables in their decision-making process compared to recurrent models (TCN, RNN, LSTM).

SHAP The results of SHAP analysis are presented in Tab. 4.

Table 4. Average impact of features on SHAP prediction for different deep models.

Feature	TFT	KAN	TCN	RNN	LSTM
PM	0.166	0.172	0.100	0.102	0.101
Wind speed	0.002	0.002	0.008	0.007	0.008
Wind direction	0.003	0.004	0.019	0.022	0.021
Temperature	0.002	0.002	0.019	0.011	0.016

Analysis of the average influence of features using the SHAP method revealed that the PM feature has the largest contribution to the predictions of all models, with average influence values ranging from 0.100 (TCN) to 0.172 (KAN). High values for TFT and KAN (0.166 and 0.172) suggest that these models rely heavily on PM, potentially at the expense of other features. Among the meteorological features, wind direction showed a larger influence in the recursive models (TCN: 0.019; RNN: 0.022; LSTM: 0.022) than in TFT (0.003) and KAN (0.004). A similar result is observed in the case of temperature. SHAP results confirm the dominance of PM in all models, while also highlighting the greater sensitivity of recursive models to meteorological features, which may be due to their ability to capture temporal dependencies in the data.

5.3 Summary

The results indicate that including meteorological variables yields a slight improvement in forecasting accuracy compared to using historical PM data alone. The proposed models (TFT, KAN, and TCN) consistently outperform the baseline RNN and LSTM approaches across all stations and evaluation metrics. The explainability analysis based on LIME and SHAP suggests that meteorological variables contribute to the predictions, although their impact remains secondary to that of historical PM concentrations. These observations confirm the robustness of the proposed architectures while indicating that meteorological information provides only limited additional benefit in short-term PM forecasting.

6 Conclusion

This study evaluated three modern neural network models: the temporal fusion transformer, the Kolmogorov-Arnold Network, and the temporal convolutional network for short-term forecasting of particulate matter concentrations using data from air quality monitoring stations located in diverse regions of Poland. The proposed models consistently achieved better forecasting accuracy than the baseline methods on all stations. Training with meteorological variables led to slight improvements over using historical PM data alone, indicating that short-term PM dynamics are dominated by their temporal structure. The use of LIME and SHAP provided additional insight into the contribution of individual meteorological factors to the prediction results.

Future research will focus on extending the models to multi-step forecasting and utilizing additional explanatory variables, such as traffic intensity or emission indicators. Another promising direction is the use of spatiotemporal approaches that exploit dependencies between neighboring stations. Further work may also explore alternative explainability techniques and their consistency with atmospheric knowledge.

Disclosure of Interests. The authors declare that they have no competing interests.

References

1. Abdullah, S., Napi, N.N.L.M., Ahmed, A.N., Mansor, W.N.W., Mansor, A.A., Ismail, M., Abdullah, A.M., Ramly, Z.T.A.: Development of multiple linear regression for particulate matter (pm10) forecasting during episodic transboundary haze event in malaysia. Atmosphere **11**(3), 289 (2020)
2. Bai, W., Suzuki, H., Qin, C., Tarroni, G., Oktay, O., Matthews, P.M., Rueckert, D.: Recurrent neural networks for aortic image sequence segmentation with sparse annotations. In: International conference on medical image computing and computer-assisted intervention. pp. 586–594. Springer (2018)
3. Bhatti, U.A., Yan, Y., Zhou, M., Ali, S., Hussain, A., Qingsong, H., Yu, Z., Yuan, L.: Time series analysis and forecasting of air pollution particulate matter (pm 2.5): an sarima and factor analysis approach. Ieee Access **9**, 41019–41031 (2021)
4. Chaves, M.G.D., Da Silva, A.B., Mercuri, E.G.F., Noe, S.M.: Particulate matter forecast and prediction in curitiba using machine learning. Frontiers in Big Data **7**, 1412837 (2024)
5. Chen, J., Zhu, S., Wang, P., Zheng, Z., Shi, S., Li, X., Xu, C., Yu, K., Chen, R., Kan, H., et al.: Predicting particulate matter, nitrogen dioxide, and ozone across great britain with high spatiotemporal resolution based on random forest models. Science of The Total Environment **926**, 171831 (2024)
6. Eren, B., Aksangür, İ., Erden, C.: Predicting next hour fine particulate matter (pm2. 5) in the istanbul metropolitan city using deep learning algorithms with time windowing strategy. Urban Climate **48**, 101418 (2023)
7. Harishkumar, K., Yogesh, K., et al.: Forecasting air pollution particulate matter (pm2. 5) using machine learning regression models. Procedia Computer Science **171**, 2057–2066 (2020)
8. Hong, W.Y.: Meteorological variability and predictive forecasting of atmospheric particulate pollution. Scientific Reports **14**(1), 14 (2024)
9. Kim, B.Y., Lim, Y.K., Cha, J.W.: Short-term prediction of particulate matter (pm10 and pm2. 5) in seoul, south korea using tree-based machine learning algorithms. Atmospheric Pollution Research **13**(10), 101547 (2022)
10. Kouziokas, G.N.: Svm kernel based on particle swarm optimized vector and bayesian optimized svm in atmospheric particulate matter forecasting. Applied Soft Computing **93**, 106410 (2020)
11. Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T.Y., Tegmark, M.: Kan: Kolmogorov-arnold networks. arXiv preprint arXiv:2404.19756 (2024)
12. Ma, X., Shen, W.: Generalized de boor–cox formulas and pyramids for multi-degree spline basis functions. Mathematics **11**(2), 367 (2023)
13. Mosca, E., Szigeti, F., Tragianni, S., Gallagher, D., Groh, G.: Shap-based explanation methods: a review for nlp interpretability. In: Proceedings of the 29th international conference on computational linguistics. pp. 4593–4603 (2022)
14. Neto, P.S.D.M., Firmino, P.R.A., Siqueira, H., Tadano, Y.D.S., Alves, T.A., De Oliveira, J.F.L., Marinho, M.H.D.N., Madeiro, F.: Neural-based ensembles for particulate matter forecasting. Ieee Access **9**, 14470–14490 (2021)
15. Pandya, A., Nanavaty, R., Pipariya, K., Shah, M.: A comparative and systematic study of machine learning (ml) approaches for particulate matter (pm) prediction. Archives of computational methods in engineering **31**(2), 595–614 (2024)
16. Ramentol, E., Grimm, S., Stinzendörfer, M., Wagner, A.: Short-term air pollution forecasting using embeddings in neural networks. Atmosphere **14**(2), 298 (2023)

17. Ren, Y., Wang, S., Xia, B.: Deep learning coupled model based on tcn-lstm for particulate matter concentration prediction. Atmospheric Pollution Research **14**(4), 101703 (2023)
18. Udristioiu, M.T., Mghouchi, Y.E., Yildizhan, H.: Prediction, modelling, and forecasting of pm and aqi using hybrid machine learning. Journal of Cleaner Production **421**, 138496 (2023)
19. Yang, C., Kai, Z., Wang, X., Hu, T., Liu, H.: An optimized tcn–lstm model for predicting pm 2.5 in metro systems. Environmental Science: Processes & Impacts (2025)
20. Zafar, M.R., Khan, N.: Deterministic local interpretable model-agnostic explanations for stable explainability. Machine Learning and Knowledge Extraction **3**(3), 525–541 (2021)
21. Zaini, N., Ahmed, A.N., Ean, L.W., Chow, M.F., Malek, M.A.: Forecasting of fine particulate matter based on lstm and optimization algorithm. Journal of Cleaner Production **427**, 139233 (2023)
22. Zalasiński, M., Cader, A., Patora-Wysocka, Z., Xiao, M.: Evaluating neural network models for predicting dynamic signature signals. Journal of Artificial Intelligence and Soft Computing Research **14** (2024)
23. Zhao, X., Song, M., Liu, A., Wang, Y., Wang, T., Cao, J.: Data-driven temporal-spatial model for the prediction of aqi in nanjing. Journal of Artificial Intelligence and Soft Computing Research **10**(4), 255–270 (2020)

A.7 Publikacja konferencyjna A6 – Potwierdzenie akceptacji artykułu – Załącznik 7

Od: **ICAISC** <icaisc@pcz.pl>
Date: wt., 28 kwi 2026 o 09:03
Subject: [ZEWNETRZNE] ICAISC Decision Notification
To: Jarosław Bernacki <jaroslaw.bernacki@pwr.edu.pl>

Dear Author(s),

On behalf of the program committee, we are pleased to inform you that your paper:

Paper ID: 12292
Title: Forecasting Particulate Matter with Neural Networks: Accuracy and Explainability using Meteorological Data

has been accepted for presentation at the 25th International Conference on Artificial Intelligence and Soft Computing 2026, and publication in the Springer LNAI conference proceedings.
We are pleased to offer you a reduced fee 5500 PLN (on site attendance),see https://www.icaisc.eu/Payment#payment. The payment can be effected by a wire transfer or by a credit card (a credit card form and our bank account number are available on our web page in the Payment Section). All costs associated to bank transfer payments must be borne by the customer.
Please provide your paper number on the bank transfer form.
In case of submission-related problems, please contact Marcin Korytkowski (marcin.korytkowski@pcz.pl).

Should you have any technical questions during paper submission or registration process please send an email to marcin.korytkowski@pcz.pl.
Should you have any other questions, please contact the conference office at icaisc@icaisc.eu or Rafał Scherer at rafal.scherer@pcz.pl; for urgent matters only, you may also call or WhatsApp him at +48 601 365 426.
We look forward to seeing you on June 14-18, 2026.

Sincerely Yours,
Leszek Rutkowski, ICAISC General Chairman

A.8 Publikacja konferencyjna A6 – Potwierdzenie uczestnictwa w konferencji – Załącznik 8

POLISH NEURAL NETWORKS SOCIETY
al. Armii Krajowej 36
42-200 Czestochowa, Poland
Phone/Fax: 48 34 3250546

June 16, 2026, Poland

This is to confirm that Jarosław Bernacki from Wrocław University of Science and Technology, Poland, participated in the 25th International Conference on Artificial Intelligence and Soft Computing ICAISC 2026 which was held in June 14-18, 2026 in Zakopane, Poland (http://icaisc.eu), and presented a paper:

Forecasting Particulate Matter with Neural Networks: Accuracy and Explainability using Meteorological Data, by:

Jarosław Bernacki
Izabela Sówka

Polskie Towarzystwo Sieci Neuronowych
42-200 Częstochowa, Al. Armii Krajowej 36
tel. (0-34)250-546

Prof. Leszek Rutkowski, General Chair

A.9 Artykuł A7 – Załącznik 9

FORECASTING PARTICULATE MATTER CONCENTRATIONS WITH DEEP NEURAL NETWORKS USING METEOROLOGICAL DATA

Jarosław Bernacki[1]
Izabela Sówka[2]
[1]Faculty of Information and Communication Technology, Wrocław University of Science and Technology, Wrocław, Poland
[2]Faculty of Environmental Engineering, Wrocław University of Science and Technology, Wrocław, Poland
Correspondence: jaroslaw.bernacki@pwr.edu.pl

Abstract

Particulate matter (PM), especially PM10 and PM2.5, poses a significant threat to human health and the environment. Accurate forecasting of particulate matter concentrations is essential for air quality management, public health protection, and the implementation of effective pollution reduction strategies. In this study, we investigate the problem of short-term forecasting of PM10 and PM2.5 concentrations using historical measurements. To address this challenge, we propose and evaluate two independent deep learning approaches: the Kolmogorov–Arnold Network (KAN) and the temporal fusion transformer (TFT). The KAN model is designed to capture complex nonlinear relationships within air quality time series, whereas the TFT architecture utilises attention mechanisms to model temporal dependencies and identify relevant patterns over time. The proposed models are trained and tested on real-world air quality datasets in two ways: using only historical concentrations and using both historical concentrations and meteorological data. The forecasting performance is assessed using standard metrics, including MAE, MSE, and R2 measures. Experimental results confirm the efficacy of the proposed methods, compared with the LSTM-based baseline model.

Keywords: air pollution, forecasting, deep models, particulate matter

Introduction

Particulate matter (PM) is one of the most harmful air pollutants due to its adverse effects on human health and the environment. Depending on particle size, particulate matter is commonly classified as PM_{10}, consisting of particles with diameters smaller than 10 μm, and $PM_{2.5}$, which includes finer ones with diameters below 2.5 μm. Exposure to elevated concentrations of PM has been associated with

DOI: 10.5604/01.3001.0055.8185
Received: 10.06.2026 Accepted: 26.06.2026 Published: 29.06.2026

numerous health problems, including respiratory diseases, cardiovascular disorders, and increased mortality rates. In addition to its impact on public health, PM contributes to environmental degradation, reduced visibility, and climate-related effects. Consequently, continuous monitoring of PM concentrations has become a critical component of air quality management systems. Accurate forecasting of PM concentrations enables authorities and citizens to take preventive actions before pollution episodes occur, for example, by reducing exposure risks and supporting informed decision-making. As a result, the development of reliable forecasting models for PM concentrations has attracted considerable attention from researchers in recent years (Kujawska et al., 2022; Liu et al., 2025).

Despite significant improvements in air quality over the past decade, PM pollution remains a serious environmental challenge in Poland (Badyda et al., 2016). Elevated concentrations of PM10 and PM2.5 are frequently observed, particularly during the heating season, when emissions from residential solid-fuel combustion increase substantially. Additional sources of PM include road transport, industrial activities, and unfavourable meteorological conditions that limit pollutant dispersion (Majewski et al., 2024). Although numerous air quality improvement programs have been implemented, exceedances of recommended PM concentration limits still occur in many urban and suburban areas. Therefore, accurate monitoring and forecasting of PM concentrations remain essential to support environmental policies and protect public health in Poland (Penkała et al., 2023; Połednik, 2022). In Poland, the air quality is monitored by the Chief Inspectorate of Environmental Protection (GIOS)[1], which also provides data online. Numerous studies have explored the prediction of air pollutant concentrations using a wide range of approaches, including machine learning methods (Gryech et al., 2024), deep learning architectures (Swetha et al., 2024), and hybrid frameworks that combine multiple modelling techniques (Borah et al., 2024; Sówka et al., 2019).

This study investigates short-term forecasting of PM_{10} and $PM_{2.5}$ concentrations using two modern neural network architectures: the Kolmogorov–Arnold Network (KAN) and the temporal fusion transformer (TFT). To evaluate their predictive capabilities, two experimental scenarios are considered. In the first scenario, the models are trained exclusively on historical PM concentrations. In the second scenario, historical PM data are supplemented with selected meteorological variables, including temperature, wind speed, wind direction, and atmospheric pressure. This experimental design enables the assessment of both the forecasting performance of KAN and TFT and the influence of meteorological information on prediction accuracy. The models are compared against an LSTM-based baseline using real-world data collected at several Polish monitoring stations that represent different environmental conditions.

[1] https://www.gov.pl/web/gios-en

Selected Previous Studies on Particulate Matter Forecasting and Meteorological Determinants

Previous studies on particulate matter forecasting have applied a broad range of statistical, machine learning, deep learning, and hybrid approaches, often incorporating meteorological variables as additional predictors. These works provide a reference background for assessing whether advanced neural network architectures can improve PM_{10} and $PM_{2.5}$ forecasting accuracy under different input configurations.

Among studies addressing the role of meteorological factors in particulate matter variability, Won et al. (2021) investigated the influence of PM hygroscopicity on atmospheric visibility. Their research focused on both coarse particulate matter (CPM) and $PM_{2.5}$, examining how meteorological conditions affect particle behaviour and optical properties. To quantify the relationship between PM concentrations and weather variables, the authors proposed a censored regression model to estimate hygroscopic growth characteristics. The approach was validated using $PM_{2.5}$ measurements collected by low-cost sensors deployed in two different regions. The findings revealed that increased concentrations of CPM and $PM_{2.5}$ reduce visibility, while relative humidity amplifies these effects. Furthermore, using humidity-dependent correction factors significantly enhanced the accuracy of low-cost sensor measurements. The study emphasises the importance of considering interactions between meteorological conditions and PM in air quality assessment and forecasting applications.

Rogula-Kozłowska et al. (2014) investigated the seasonal variability of $PM_{2.5}$ concentrations and their chemical composition at three monitoring sites located in different regions of Poland: Katowice, Gdańsk, and Diabla Góra. Their analysis revealed elevated $PM_{2.5}$ levels at all locations, with the highest concentrations observed during winter in Katowice. Carbonaceous species, secondary inorganic aerosols, and benzo(a)pyrene were identified as the predominant constituents, with the latter contributing substantially to increased health risks during the heating season.

Vovk et al. (2024) introduced a hybrid modelling framework for $PM_{2.5}$ estimation in Poland that combines the EMEP4PL chemical transport model with a random forest algorithm. The proposed approach uses information from the transport model together with meteorological variables, emission data, land-use characteristics, and temporal predictors. The proposed solution achieved improved predictive performance, including higher coefficients of determination, lower estimation bias, and enhanced capability to identify high-pollution events, compared with the standalone chemical transport model.

Kowalski et al. (2020) explored the application of advanced neural network techniques for forecasting PM_{10} concentrations over a 24-hour horizon. The authors assessed the proposed method using multiple error metrics and benchmarked its performance against alternative forecasting approaches. Their results demonstrated that convolutional neural networks can provide accurate short-term air quality predictions and are a promising tool for PM forecasting.

Czernecki et al. (2021) investigated the use of machine learning techniques for short-term forecasting of PM_{10} and $PM_{2.5}$ concentrations in four major Polish urban agglomerations, utilising a decade of observational data from background, traffic, and industrial monitoring stations. The study compared several models, including stepwise regression, random forest, XGBoost, and neural networks, and found that XGBoost achieved the best predictive performance, whereas stepwise regression yielded the weakest results.

Tong et al. (2024) proposed a transformer-based framework for spatiotemporal prediction of $PM_{2.5}$ concentrations, using aerosol optical depth data collected over California. The developed model demonstrated improved performance over LSTM and other baseline methods across multiple evaluation metrics.

Mauricio et al. (2024) explored the application of transformer architectures for forecasting CPM in Mexico City across different prediction horizons. Their results indicate that the transformer model outperforms traditional approaches, such as ARIMA and LSTM across multiple monitoring sites; however, the study also notes limitations in short-term prediction accuracy, particularly in areas characterised by strong spatial variability and proximity to industrial emission sources.

Cichowicz et al. (2020) analysed PM_{10} concentrations in Poland during winter seasons, showing that low wind speeds significantly contribute to increased air pollution levels. Based on data from three monitoring stations, the study identified smog episodes (the so-called black smog) and validated measurements from a site located near a large waste incineration facility. The results further confirm the important role of wind direction in pollutant transport and local accumulation.

Tariq et al. (2023) addressed the problem of forecasting $PM_{2.5}$ concentrations in underground subway environments. A hybrid framework that combines a gated probabilistic transformer with a genetic algorithm-based quantile scheduling mechanism is proposed to enhance early-warning capabilities and support ventilation control. The model provides more accurate sequential risk predictions, tighter uncertainty bounds, and improved control of $PM_{2.5}$ levels, compared to conventional approaches.

Ramentol et al. (2023) focused on predicting NO_2 concentrations using machine learning techniques based on historical pollutant and meteorological data, with the city of Erfurt (Germany) as the study area. The model utilises embedded temporal features that allow for capturing local influences, such as traffic patterns and special events. The experimental results demonstrate that the proposed approach outperforms competing methods, particularly during periods when standard seasonal patterns are disrupted, such as holidays or irregular emission conditions.

Sharma et al. (2020) proposed a hybrid deep learning framework that integrates a convolutional neural network with an LSTM architecture to forecast hourly total suspended particle (TSP) concentrations in Australia. The developed model was extensively evaluated and demonstrated improved performance compared to an ensemble of five alternative machine learning approaches.

Kujawska et al. (2022) compared several machine learning techniques for PM_{10} concentration forecasting in Lublin, including both meteorological and chemical variables. Among the tested models, the artificial neural network achieved the best predictive performance, effectively identifying exceedance risks of PM_{10} levels.

Kryza et al. (2019) investigated air quality forecasting in Poland using the WRF-Chem model and compared two emission estimation approaches: the baseline GENEMIS method and the heating degree day (HDD) approach. The results indicated that both methods produced comparable performance. However, the HDD approach better captured pollutant concentrations during warmer periods, while the baseline method tended to overestimate values.

Du et al. (2022) proposed a hybrid $PM_{2.5}$ forecasting and impact assessment framework that integrates data preprocessing, feature extraction, and an optimised extreme learning machine (ELM) model tuned using the Harris hawk optimisation algorithm. In addition to accurate $PM_{2.5}$ prediction, the system estimates associated health impacts and economic losses. Experiments conducted on datasets from Beijing, Tianjin, and Shijiazhuang demonstrated the effectiveness of the proposed approach for supporting environmental management, emission-reduction strategies, and public health protection.

The selected studies indicate that forecasting performance in particulate matter prediction depends on both the modelling approach and the type of input data used. In particular, the inclusion of meteorological variables has been shown to improve the representation of short-term PM variability under different environmental conditions. Against this background, the present study evaluates KAN and TFT architectures for PM_{10} and $PM_{2.5}$ forecasting under two input configurations: using historical PM concentrations alone and in combination with selected meteorological variables, with LSTM adopted as the baseline reference model.

Methodology and Experimental Design

Data Sources and Study Area

The performance of the proposed forecasting approaches was assessed through simulation experiments using real-world observations from the Chief Inspectorate of Environmental Protection (GIOS). The dataset contains historical concentration values (µg/m³) for multiple air pollutants. The main objective is to train the developed models on PM_{10} and $PM_{2.5}$ time series collected from several Polish cities (referred to as the original dataset) and using selected meteorological variables. The predicted concentrations are subsequently compared with the observed values to evaluate forecasting accuracy. The dataset covers the period from January 2022 to December 2024.

The meteorological inputs include temperature (°C), wind speed (km/h), wind direction (°), and atmospheric pressure (hPa). These variables were obtained from the Meteostat platform, which aggregates historical weather data from multiple global sources. The meteorological dataset corresponds to the same time interval as the air quality measurements, i.e., from January 2022 to December 2024.

The selected features were chosen for their completeness and relevance to PM dynamics. Temperature influences chemical reaction rates in the atmosphere, wind conditions affect pollutant dispersion, and atmospheric pressure contributes to the formation of inversion layers that can trap pollutants near the surface and increase their concentrations. In this study, temperature is represented by the daily mean, ensuring consistency with the dataset. For this study, the dataset provided a consistent set of observations, including both PM concentrations and meteorological variables required for model training and evaluation.

The evaluation was based on historical concentrations of PM_{10} and $PM_{2.5}$ measured at the air quality monitoring stations in Poland. The locations of the monitoring stations are reported with geographic coordinates (latitude φ, longitude λ) in the WGS84 reference system. The analysed locations are the following:

- Zielonka Bory Tucholskie (KpZielBoryTu; 53,662117° N; 17,934017° E);
- Zakopane (MpZakopaSien; 49,2936° N; 19,9601° E);
- Częstochowa – SlCzestoZana (50,801918° N; 19,106961° E – PM2.5) / SlCzestoBacz (50,836389° N; 19,130111° E – PM10);
- Gliwice/Knurów – SlGliwicMewy (50,279481° N; 18,655736° E – PM2.5) / SlKnurJedNar (50,233167° N; 18,655722° E – PM10);
- Szczecin (ZpSzczec1Maj; 53,712114° N; 16,692517° E).

In the Częstochowa and Gliwice/Knurów regions, PM_{10} and $PM_{2.5}$ concentrations are measured at multiple monitoring stations located in close spatial proximity. For example, the stations SlCzestoZana and SlCzestoBacz are positioned near each other in Częstochowa, while SlGliwicMewy and SlKnurJedNar form a similar pair in the Gliwice–Knurów area. Although the relatively short distances between these stations suggest comparable measurement conditions, even minor spatial separations (e.g., on the order of several hundred meters) can still result in noticeable differences in recorded particulate matter levels due to local emission sources and microenvironmental conditions. The geographical locations of the monitoring stations considered are illustrated in Figure 1.

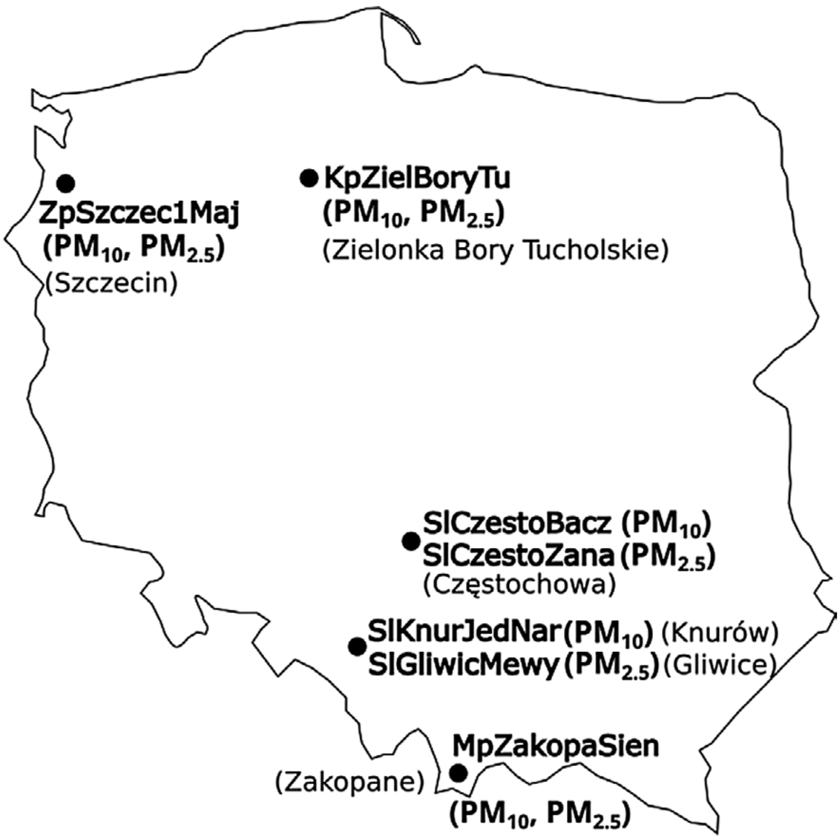


Figure 1. The location of the air quality monitoring stations in Poland used in the experiment

The selection of monitoring stations was based primarily on the completeness and temporal continuity of the available datasets over the considered study period, ensuring the absence of significant gaps in both pollutant and meteorological observations. Furthermore, the chosen stations were selected to represent a range of environmental conditions across Poland. Specifically, Zielonka Bory Tucholskie (KpZielBoryTu) represents rural background conditions, Zakopane (MpZakopaSien) corresponds to a mountainous environment, while Częstochowa (SlCzestoBacz / SlCzestoZana), Gliwice/Knurów (SlGliwicMewy / SlKnurJedNar), and Szczecin (ZpSzczec1Maj) reflect urban and industrial settings. This diversity of locations enables a more comprehensive evaluation of the proposed models under varying air quality regimes and contributes to assessing their robustness and generalisation capability.

Input Variables and Forecasting Configurations

The forecasting task was defined as a next-day prediction of PM_{10} or $PM_{2.5}$ concentration based on historical observations. Two input configurations were considered. In the first configuration, only historical PM concentrations were used as predictors. In the second configuration, historical PM concentrations were supplemented with selected meteorological variables, including temperature, wind speed, wind direction, and atmospheric pressure.

For the full input configuration, the input vector at time step *t* included PM concentration and meteorological variables and was defined as follows:

$$x_t = [x^{conc}, x^{temp}, x^{wind}, x^{wdir}, x^{pres}]$$

where x^{conc} denotes historical PM concentration, x^{temp} denotes temperature in °C, x^{wind} denotes wind speed in km/h, x^{wdir} denotes wind direction in degrees, and x^{pres} denotes atmospheric pressure in hPa. The model output was a single forecasted value of PM_{10} or $PM_{2.5}$ concentration for the next time step.

Data Preprocessing and Training Procedure

The input data were prepared using a sliding window approach. For each sample, five consecutive daily observations were used to predict the PM_{10} or $PM_{2.5}$ concentration for the following day. The five-day input window was selected to capture short-term temporal dependencies while maintaining a sufficient number of training samples. Shorter sequences may not fully represent relevant temporal patterns, whereas longer windows may reduce the number of available training examples and increase the risk of overfitting.

The dataset was split chronologically into training, validation, and testing subsets using a 75/15/15 ratio. The first 75% of the time series was used for model training, the next 15% for validation, and the remaining 15% for independent evaluation. All input features, except wind direction, were normalised to the [0, 1] range. Wind direction was transformed using circular encoding, in which the angle θ was represented by its sine and cosine components. This procedure avoids discontinuities associated with angular variables and prevents the artificial separation of values near 0° and 360°.

The hyperparameters of the models were determined empirically. Training was performed using the Adam optimiser with a learning rate of 0.001 and the mean squared error (MSE) loss function. To reduce the risk of overfitting, early stopping was applied when no improvement in validation loss was observed. In practice, convergence typically occurred after approximately 1000 epochs.

The forecasting models and baseline method were implemented in Python using the PyTorch framework with CUDA acceleration. Statistical analyses were carried out in MATLAB. All experiments were performed on a Gigabyte Aero laptop equipped with an Intel Core i7-13700H processor, 32 GB of RAM, and NVIDIA GeForce RTX 4070 GPU with 8 GB of VRAM.

Forecasting Models

The forecasting framework included two deep neural network architectures: the Kolmogorov–Arnold Network (KAN) and the temporal fusion transformer (TFT). Both models were trained and evaluated using the same input configurations, preprocessing procedure, and the train-test split. The architectures of the considered models are presented below.

Kolmogorov–Arnold Network

Kolmogorov–Arnold Networks (KAN) are motivated by the Kolmogorov–Arnold representation theorem, which states that any multivariate continuous function can be decomposed into a composition of univariate continuous functions. This theoretical result forms the basis of the KAN architecture. KAN explicitly parameterises nonlinear transformations with learnable one-dimensional functions, thereby enhancing its ability to approximate complex nonlinear relationships.

In contrast to conventional neural networks that rely on fixed activation functions, the KAN architecture employs learnable univariate functions defined on network connections. These functions are represented using B-spline basis functions, whose values are computed through the Cox–de Boor recursion formula (Ma & Shen, 2023). During training, the spline coefficients are optimised together with the network parameters, allowing the model to adaptively learn nonlinear transformations from data. In this study, the adopted KAN model consists of multiple KAN layers with 128, 256, 128, and 64 hidden units, followed by a linear output layer producing the PM concentration forecast. Dropout regularisation is applied between successive layers to reduce overfitting. The illustrative scheme of the discussed KAN architecture is presented in Figure 2.

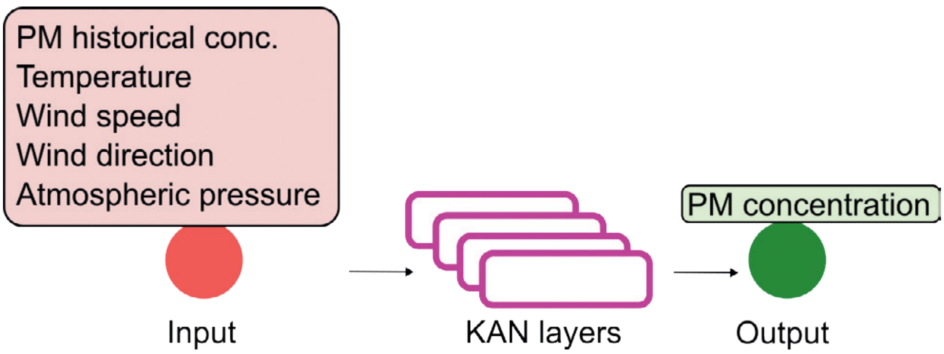


Figure 2. The structure of the KAN network

The input data are processed using a sliding-window approach, in which each sample includes historical PM measurements and, when available, meteorological variables. These windows are flattened and fed into the network, allowing it to capture nonlinear temporal dependencies without relying on recurrent or convolutional structures. The size of the input window is treated as a hyperparameter and optimised during training.

Temporal Fusion Transformer

The temporal fusion transformer (TFT) is a hybrid deep learning architecture that integrates recurrent neural networks with an attention mechanism, enabling the model to capture both short-term and long-term temporal dependencies while simultaneously assessing the relevance of input variables.

Each input feature, including PM concentrations and meteorological variables, is first transformed into a unified representation through embedding and normalisation layers, ensuring consistent scaling and effective integration of heterogeneous data sources.

A variable selection network, built on gated residual networks, is applied at each time step to dynamically assign importance weights to individual predictors. This mechanism allows the model to adaptively determine which variables have the strongest influence on the forecast at a given moment.

The temporal processing component consists of LSTM layers: two in the encoder and one in the decoder, each with 128 hidden units. The encoder extracts short- and medium-term temporal patterns from historical sequences, while the decoder generates representations for the prediction horizon, preserving temporal coherence between past observations and future outputs.

To further enhance temporal modelling, a multi-head self-attention mechanism is applied after the LSTM block. This module captures dependencies across all time steps, enabling the identification of critical historical events (e.g., pollution peaks) and the learning of long-range temporal relationships. The multi-head self-attention module employs four attention heads operating on a hidden representation of dimension 128. Furthermore, two transformer layers are used to model long-range temporal dependencies and interactions between different time steps in the input sequence. The attention mechanism is defined as:

$$Attention(Q, K, V) = softmax(QK^T) \cdot V$$

which computes adaptive weights reflecting the importance of each time step in the forecasting process.

In the attention mechanism, **Q** (Query), **K** (Key), and **V** (Value) are learned representations derived from the input sequence. The query matrix **Q** represents the information for which the model seeks relevant context, while the key matrix

K contains descriptors used to measure the similarity between different time steps. The value matrix **V** stores aggregated information passed to subsequent layers. The product ($\mathbf{QK}^T$) computes similarity scores between queries and keys, and the softmax function converts these scores into normalised attention weights. These weights are then applied to **V**, allowing the model to focus on the most relevant observations when generating a forecast. Finally, the aggregated representations are passed through a fully connected layer to produce PM_{10} and $PM_{2.5}$ forecasts. The model supports both single-step and multi-step prediction horizons, allowing estimation of future concentrations over (t+1) or (t+1, …, t+H).

The scheme of the discussed approach is presented in Figure 3.

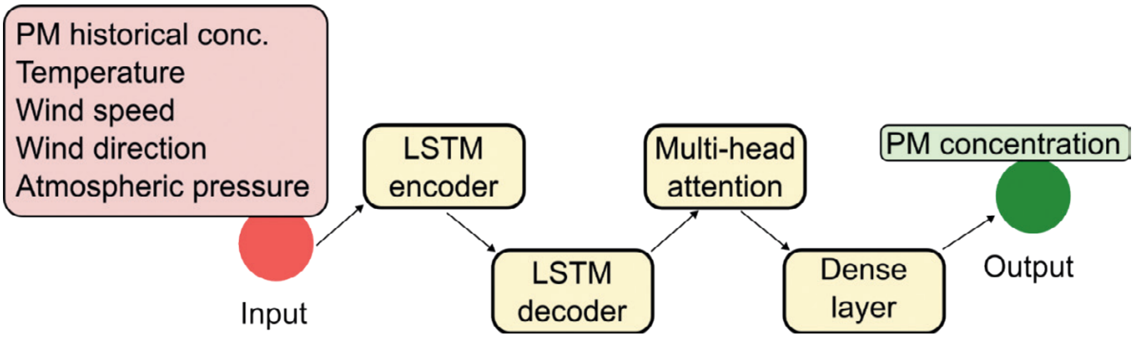


Figure 3. The structure of the TFT framework

Baseline Model

A long short-term memory (LSTM) network was used as a baseline model to serve as a reference for evaluating the performance of the proposed KAN and TFT architectures. The LSTM model was trained and tested under the same two input configurations as the proposed models: using only historical PM concentrations and using historical PM concentrations supplemented with meteorological variables.

The baseline network consisted of four LSTM layers with 64 hidden units, followed by a fully connected linear layer transforming the hidden representation into a single forecasted PM_{10} or $PM_{2.5}$ concentration value. The same training procedure and hyperparameter settings were applied to the LSTM model as to the proposed architectures, enabling a consistent comparison of forecasting performance.

Evaluation Metrics

The forecasting performance of the proposed KAN and TFT models and the baseline LSTM network was evaluated using two standard error metrics: mean absolute error (MAE, eq. 1), mean squared error (MSE, eq. 2), and coefficient of determination (R^2, eq. 3). These metrics were calculated by comparing the predicted PM10 and PM2.5 concentrations with the corresponding observed values in the test dataset.

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |x_t - y_t| \tag{1}$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n} (x_t - y_t)^2 \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (x_t - y_t)^2}{\sum_{i=1}^{n} (x_t - y_{mean})^2} \tag{3}$$

where x_t is the actual value, y_t is a predicted value, y_{mean} is the average real value, and n is the number of input observations. Lower MAE and MSE values indicate better forecasting performance. Since PM concentration values were normalised to the [0,1] min-max range before model training, the reported MAE, MSE, and R^2 values refer to the normalised scale rather than to the original concentration units expressed in $\mu g/m^3$. Therefore, these metrics should be interpreted primarily as comparative measures of forecasting performance between models, input configurations, and monitoring stations.

Results and Discussion

The results for $PM_{2.5}$ forecasting experiments are presented in Table 1.

The obtained results indicate that both proposed architectures, KAN and TFT, consistently outperform the baseline LSTM model across all considered monitoring stations and evaluation metrics. In terms of MAE, proposed models achieve the lowest error values, typically ranging between 0.11 and 0.12, while LSTM shows noticeably higher errors, reaching up to 0.17–0.20 depending on the station. A comparison between models trained with and without meteorological data (denoted by the asterisk) demonstrates a lower forecasting accuracy when meteorological variables are excluded. For both KAN* and TFT*, higher MAE and MSE values are observed at all locations, confirming that the inclusion of meteorological information improves forecasting accuracy. This effect is particularly evident in more variable environments, such as Szczecin and Gliwice, where the differences between full and reduced input configurations are more visible. In terms of MSE, KAN generally achieves the most stable and lowest error values through stations, closely followed by TFT, while LSTM exhibits the highest variability and error levels. Although in some cases TFT shows competitive or slightly lower MSE values, its performance remains consistently better when meteorological inputs are included. For all monitoring stations and forecasting models, higher R^2 values were obtained when meteorological variables were included, indicating that weather-related information improves the models' ability to explain the variability of PM concentrations. MAE results are summarised in Figure 4.

Table 1. Error measure values for the proposed methods from all the air quality monitoring stations for the $PM_{2.5}$ fraction

Air station	Error measure	KAN	KAN*	TFT	TFT*	LSTM	LSTM *
Zielonka Bory Tucholskie	MAE	0.13	0.11	0.14	0.11	0.17	0.16
	MSE	0.04	0.03	0.06	0.04	0.07	0.06
	R^2	0.87	0.90	0.87	0.89	0.85	0.86
Zakopane	MAE	0.12	0.11	0.15	0.12	0.20	0.15
	MSE	0.05	0.04	0.04	0.03	0.08	0.05
	R^2	0.86	0.89	0.87	0.91	0.84	0.85
Częstochowa	MAE	0.11	0.11	0.15	0.11	0.18	0.17
	MSE	0.05	0.03	0.05	0.03	0.07	0.05
	R^2	0.87	0.89	0.87	0.90	0.85	0.86
Gliwice	MAE	0.14	0.12	0.15	0.11	0.18	0.17
	MSE	0.05	0.04	0.04	0.03	0.06	0.06
	R^2	0.86	0.89	0.87	0.89	0.84	0.86
Szczecin	MAE	0.14	0.11	0.15	0.12	0.19	0.15
	MSE	0.04	0.03	0.05	0.03	0.08	0.06
	R^2	0.87	0.90	0.88	0.90	0.84	0.86

MAE and MSE values are reported on the normalised [0,1] scale. The KAN*, TFT*, and LSTM* denote the proposed and baseline methods trained, including the meteorological data, while methods denoted without * were trained using only historical PM data.

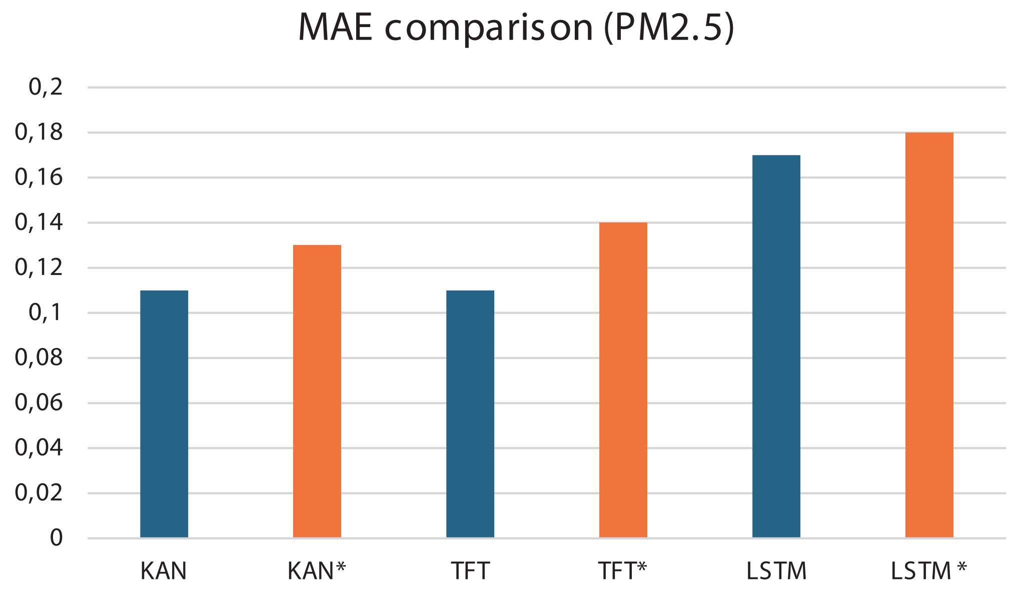


Figure 4. MAE comparison for $PM_{2.5}$ forecasting for all considered models

Overall, the results demonstrate that using meteorological variables improves prediction accuracy across all tested models, and that KAN and TFT provide more robust and accurate forecasts than the classical LSTM architecture under diverse environmental conditions. Table 2 presents the results of PM_{10} forecasting evaluation.

Table 2. Error measure values for the proposed methods from all the air quality monitoring stations for the PM_{10} fraction.

Air station	Error measure	KAN	KAN*	TFT	TFT*	LSTM	LSTM *
Zielonka Bory Tucholskie	MAE	0.12	0.10	0.14	0.11	0.15	0.14
	MSE	0.04	0.03	0.05	0.03	0.06	0.05
	R^2	0.87	0.90	0.87	0.90	0.83	0.86
Zakopane	MAE	0.13	0.11	0.15	0.12	0.19	0.15
	MSE	0.04	0.03	0.04	0.03	0.08	0.04
	R^2	0.87	0.89	0.87	0.90	0.84	0.85
Częstochowa	MAE	0.11	0.10	0.13	0.11	0.19	0.16
	MSE	0.05	0.03	0.05	0.03	0.07	0.05
	R^2	0.86	0.89	0.88	0.91	0.84	0.85
Gliwice	MAE	0.15	0.12	0.14	0.12	0.18	0.16
	MSE	0.05	0.03	0.04	0.03	0.06	0.05
	R^2	0.87	0.90	0.87	0.90	0.85	0.86
Szczecin	MAE	0.14	0.11	0.14	0.12	0.18	0.15
	MSE	0.04	0.03	0.05	0.03	0.07	0.05
	R^2	0.87	0.89	0.87	0. 90	0.83	0.86

MAE and MSE values are reported on the normalised [0,1] scale. The KAN*, TFT*, and LSTM* denote the proposed and baseline methods trained, including the meteorological data, while methods denoted without * were trained using only historical PM data.

The results confirm the overall advantage of the KAN and TFT architectures over the baseline LSTM model on all monitoring stations and evaluation metrics. KAN achieves the lowest error values in most cases (MAE is equal to 0.10–0.12), closely followed by TFT, while LSTM consistently yields higher prediction errors, reaching up to 0.19 depending on the station and configuration. A comparison between models trained on full input data and those trained without meteorological variables shows reduced forecasting performance when meteorological information is excluded. Both KAN* and TFT* exhibit higher MAE and MSE values at all locations, indicating that meteorological factors contribute positively to model accuracy. The inclusion of meteorological variables also resulted in higher R^2 values for the considered models. This improvement is particularly visible in more heterogeneous environments such as Zakopane and Gliwice, where weather conditions strongly influence pollutant dynamics. Regarding MSE, both KAN and TFT achieve stable and competitive performance, generally outperforming LSTM through all stations. In several cases, TFT achieves results comparable to KAN, although KAN remains slightly more consistent, with lower error values. MAE results are visible in Figure 5.

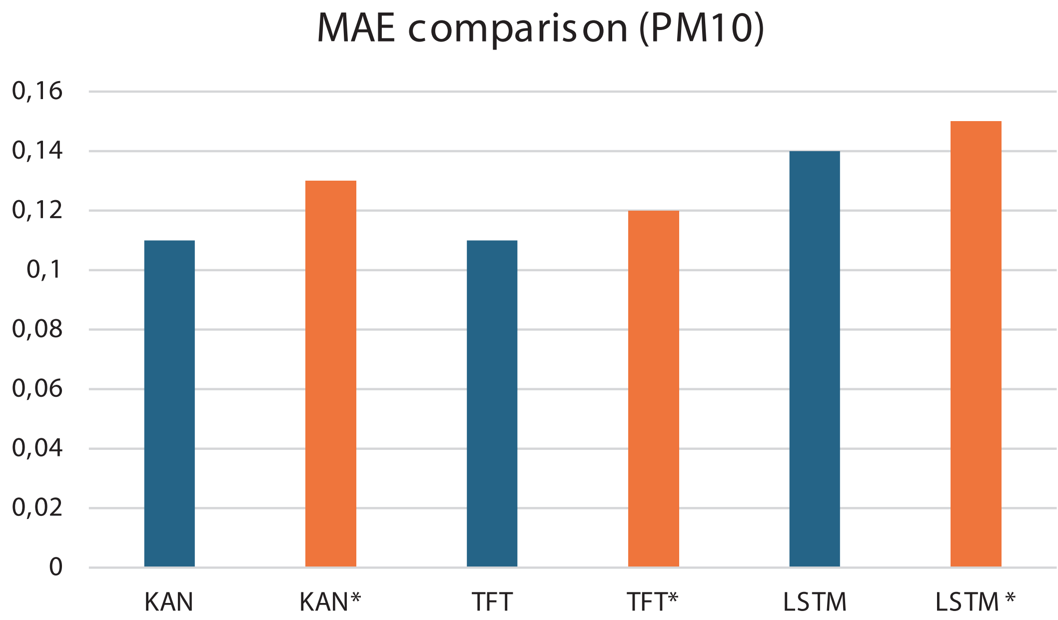


Figure 5. MAE comparison for PM_{10} forecasting for all considered models

Overall, the findings reinforce the claim that utilising meteorological variables enhances forecasting accuracy and that both KAN and TFT provide more reliable predictions than LSTM.

The results indicate the high predictive potential of the KAN and TFT architectures. However, their interpretation should take into account the scope and assumptions of the conducted analyses. Forecasting performance may be partly related to the length of the available time series, as longer datasets usually capture long-term trends, interannual variability, and less frequent episodes of elevated pollutant concentrations more comprehensively. Seasonal variability in PM_{10} and $PM_{2.5}$ concentrations is also important, particularly under Polish conditions, where heating-related emissions and meteorological conditions play an important role in shaping particulate matter levels.

The results may also have been influenced by the use of a fixed five-day input window across all experiments. This approach ensured consistency in comparing the analysed models. However, future studies should examine whether different input window lengths affect forecasting accuracy. Differences between monitoring stations may also be relevant, resulting, among others, from the activity of local emission sources, the degree of urbanisation, topographical conditions, and meteorological conditions. Therefore, despite the promising results obtained for the analysed locations, their generalisation to other areas should be made with appropriate caution and supported by further validation using data from additional monitoring stations.

Conclusion

This study proposed and evaluated two deep learning approaches, KAN and TFT, for short-term forecasting of PM_{10} and $PM_{2.5}$ concentrations using historical air quality data, with and without meteorological variables. Experiments conducted at multiple monitoring stations in Poland showed that both models consistently outperform the LSTM baseline on the MAE and MSE metrics. The results also confirm that using meteorological data improves forecasting accuracy for all tested models. Overall, KAN and TFT demonstrate strong potential for reliable air-quality prediction across diverse environmental conditions. However, the transferability of these findings should be further verified using longer time series, alternative input-window lengths, season-specific analyses, and data from additional monitoring stations.

Future work will focus on extending the proposed KAN and TFT frameworks to longer forecasting horizons and evaluating their performance under real-time operational conditions. Further improvements may include utilising additional environmental and emission-related variables, as well as exploring spatial dependencies between monitoring stations using spatiotemporal modelling techniques. Moreover, future studies could investigate the robustness of the models under missing or noisy data and assess the transferability of the proposed approaches to other countries and pollution contexts.

Funding

This research received no external funding.

References

Badyda, A., Gayer, A., Czechowski, P. O., Majewski, G., & Dąbrowiecki, P. (2016). Pulmonary Function and Incidence of Selected Respiratory Diseases Depending on the Exposure to Ambient PM_{10}. *International Journal of Molecular Sciences*, *17*(11), 1954. https://doi.org/10.3390/ijms17111954

Borah, J., Nadzir, M. S. M., & Cayetano, M. G., Majumdar, S., Ghayvat, H., & Srivastava, G. (2024). AiCareAir: Hybrid-Ensemble Internet of Things Sensing Unit Model for Air Pollutant Control. *IEEE Sensors Journal.*

Cichowicz, R., Wielgosiński, G., & Fetter, W. (2020). Effect of wind speed on the level of particulate matter PM10 concentration in atmospheric air during winter season in vicinity of large combustion plant. *Journal of Atmospheric Chemistry*, 77, 35–48.

Czernecki, B., Marosz, M., & Jędruszkiewicz, J. (2021). Assessment of Machine Learning Algorithms in Short-term Forecasting of PM_{10} and $PM_{2.5}$ Concentrations in Selected Polish Agglomerations. *Aerosol Air Qual. Res.*, 21, 200586.

Du, P., Wang, J., Yang, W., & Niu, T. (2022). A novel hybrid fine particulate matter ($PM_{2.5}$) forecasting and its further application system: Case studies in China. *Journal of Forecasting*, 41, 64–85.

Gryech, I., Asaad, C., Ghogho, M., & Kobbane, A. (2024). Applications of machine learning & Internet of Things for outdoor air pollution monitoring and prediction: A systematic literature review. *Engineering Applications of Artificial Intelligence*, 137, 109182.

Jasiński, R., Galant-Gołębiowska, M., Nowak, M., Ginter, M., Kurzawska, P., Kurtyka, K., & Maciejewska, M. (2021). Case Study of Pollution with Particulate Matter in Selected Locations of Polish Cities. *Energies*, *14*(9), 2529.

Kowalski, P., Sapała, K., & Warchałowski, W. (2020). PM_{10} Forecasting Through Applying Convolution Neural Network Techniques. *International Journal of Environmental Impacts, 3*(1), 31–43.

Kryza, M., Werner, M., & Dore, A.-J. (2019). Application of degree-day factors for residential emission estimate and air quality forecasting. *International Journal of Environment and Pollution, 65*(4), 325–336.

Kujawska, J., Kulisz, M., Oleszczuk, P., & Cel, W. (2022). Machine Learning Methods to Forecast the Concentration of PM10 in Lublin, Poland. *Energies*, *15*(17), 6428.

Liu, Y., Xu, F., Liu, W., Liu, X., & Wang, D. (2025). Characteristics, Sources, Exposure, and Health Effects of Heavy Metals in Atmospheric Particulate Matter. *Current Pollution Reports, 11*(16).

Ma, X., & Shen, W. (2023). Generalized de Boor–Cox Formulas and Pyramids for Multi-Degree Spline Basis Functions. *Mathematics, 11*(2), 367. https://doi.org/10.3390/math11020367

Majewski, G., Rogula-Kozłowska, W., Rozbicka, K. et al. (2018). Concentration, Chemical Composition and Origin of PM1: Results from the First Long-term Measurement Campaign in Warsaw (Poland). *Aerosol Air Qual. Res*., 18, 636–654. https://doi.org/10.4209/aaqr.2017.06.0221

Mauricio-Alvarez, L.-E., Aceves-Fernandez, M.-A., Pedraza-Ortega, J.-C., & Ramos-Arreguin, J.-M. (2024). Evaluation of a transformer-based model for the temporal forecast of coarse particulate matter (PMCO) concentrations. *Earth Science Informatics*, 17, 3095–3110.

Nidzgorska-Lencewicz, J. (2018). Application of Artificial Neural Networks in the Prediction of PM10 Levels in the Winter Months: A Case Study in the Tricity Agglomeration, Poland. *Atmosphere*, *9*(6), 203.

Penkała, M., Rogula-Kozłowska, W., Ogrodnik, P., Bihałowicz, J. S., & Iwanicka, N. (2023). Exploring the relationship between particulate matter emission and the construction material of road surface: Case study of highways and motorways in Poland. *Materials*, 16, 1200.

Połednik, B. (2022). Emissions of air pollution in industrial and rural region in Poland and health impacts. *Journal of Ecological Engineering*, 23, 250258.

Ramentol, E., Grimm, S., Stinzendorfer, M., & Wagner, A. (2023). Short-Term Air Pollution Forecasting Using Embeddings in Neural Networks. *Atmosphere*, 14, 298.

Rogula-Kozłowska, W., Klejnowski, K., Rogula-Kopiec, P., Ośródka, L., Krajny, E., Błaszczak, B., & Mathews, B. (2013). Spatial and seasonal variability of the mass concentration and chemical composition of $PM_{2.5}$ in Poland. *Air Quality, Atmosphere & Earth, 7*, 41–58.

Sharma, E., Deo, R.C., Prasad, R., Parisi, A.V., & Raj, N. (2020). Deep air quality forecasts: suspended particulate matter modeling with convolutional neural and long short-term memory networks. *IEEE Access*, 8, 209503–209516.

Sowka, I., Chlebowska-Styś, A., Pachurka, Ł., Rogula-Kozłowska, W., & Mathews, B. (2019). Analysis of particulate matter concentration variability and origin in selected urban areas in Poland. Sustainability 2019, 11, 5735.

Swetha, G., Datla, R., Vishnu, C., & Mohan, C. K. (2024). M2-APNet: A multimodal deep learning network to predict major air pollutants from temporal satellite images. *Journal of Applied Remote Sensing*, 18, 012005–012005.

Tariq, S., Loy-Benitez, J., & Yoo, C. (2023). Enhancing the sustainable management of fine particulate matter-related health risks at subway stations through sequential forecast and gated probabilistic transformer. *Building and Environment, 244*(October 1), 110780.

Tong, W., Limperis, J., Hamza-Lup, F., Hu, Y., & Li, L. (2024). Robust Transformer-based model for spatiotemporal $PM_{2.5}$ prediction in California. *Earth Science Informatics, 17*, 315–328.

Vovk, T., Kryza, M., & Werner, M. (2024). Using random forest to improve EMEP4PL model estimates of daily $PM_{2.5}$ in Poland. Atmospheric Environment, 332, 120615.

Won, W.-S., Oh, R., Lee, W., Ku, S., Su, P.-C., & Yoon, Y.-J. (2021). Hygroscopic properties of particulate matter and effects of their interactions with weather on visibility. *Scientific Reports*, 11, 16401.

A.10 Charakterystyka wkładu autorskiego oraz zakresu prac zrealizowanych w ramach publikacji włączonych do rozprawy – Załącznik 10

Praca A1 – J. Bernacki J., R. Scherer (2025). A Comprehensive Review of Data-Driven Techniques for Air Pollution Concentration Forecasting. Sensors 25 (19), 6044

1. Jarosław Bernacki:
   - Analiza stanu światowej literatury na temat prognozowania stężeń zanieczyszczeń powietrza z podziałem na techniki prognozowania (sieci sensoryczne, uczenie maszynowe/sieci neuronowe, metody statystyczne, hybrydowe oraz pozostałe modele) oraz przygotowanie tekstu manuskryptu wraz z opracowaniem elementów wspomagających (tabele, rysunki)
2. Rafał Scherer:
   - Analiza, edycja tekstu oraz krytyczna recenzja manuskryptu.

Praca A2 – J. Bernacki (2025). Forecasting the air pollution concentration with neural networks. Urban Climate (Elsevier); Vol. 59 102262

1. Jarosław Bernacki:
   - Analiza stanu światowej literatury na temat prognozowania stężeń zanieczyszczeń powietrza, zaprojektowanie i implementacja metod prognozowania zanieczyszczeń za pomocą sieci neuronowych z uwzględnieniem danych meteorologicznych, przeprowadzenie eksperymentów ewaluacyjnych proponowanych metod z analizą wpływu danych meteorologicznych na jakość prognoz oraz przygotowanie tekstu manuskryptu wraz z opracowaniem elementów wspomagających (tabele, rysunki).

Praca A3 – J. Bernacki (2025). Forecasting the concentration of the components of the particulate matter in Poland using neural networks. Environmental Science and Pollution Research (Springer); Vol. 32, pp 9179–9212

1. Jarosław Bernacki:
   - Analiza stanu światowej literatury na temat prognozowania stężeń zanieczyszczeń powietrza, zaprojektowanie i implementacja metod prognozowania zanieczyszczeń za pomocą sieci neuronowych z uwzględnieniem danych meteorologicznych, przeprowadzenie eksperymentów ewaluacyjnych proponowanych metod z analizą wpływu danych meteorologicznych na jakość prognoz oraz przygotowanie tekstu manuskryptu wraz z opracowaniem elementów wspomagających (tabele, rysunki).

Praca A4 – J. Bernacki (2025). A hybrid transformer and symbolic regression model for weather-dependent particulate matter forecasting in air quality management. Theoretical and Applied Climatology (Springer); Vol. 156, No. 662

1. Jarosław Bernacki:

   - Analiza stanu światowej literatury na temat prognozowania stężeń zanieczyszczeń powietrza, zaprojektowanie i implementacja metod prognozowania zanieczyszczeń za pomocą sieci neuronowych z uwzględnieniem danych meteorologicznych, przeprowadzenie eksperymentów ewaluacyjnych proponowanych metod z analizą wpływu danych meteorologicznych na jakość prognoz oraz przygotowanie tekstu manuskryptu wraz z opracowaniem elementów wspomagających (tabele, rysunki).

Praca A5 – J. Bernacki (2025). Particulate Matter Forecasting Using Hybrid Autoencoders: The Role of Meteorological Data. Annales UMCS Sectio B – Geographia, Geologia, Mineralogia et Petrographia, Vol. 80, pp 265-283

1. Jarosław Bernacki:

   - Analiza stanu światowej literatury na temat prognozowania stężeń zanieczyszczeń powietrza, zaprojektowanie i implementacja metod prognozowania zanieczyszczeń za pomocą sieci neuronowych z uwzględnieniem danych meteorologicznych, przeprowadzenie eksperymentów ewaluacyjnych proponowanych metod z analizą wpływu danych meteorologicznych na jakość prognoz oraz przygotowanie tekstu manuskryptu wraz z opracowaniem elementów wspomagających (tabele, rysunki).

Praca A6 – J. Bernacki, I. Sówka. Forecasting Particulate Matter with Neural Networks: Accuracy and Explainability using Meteorological Data. Praca zaakceptowana na konferencję ICAISC 2026

1. Jarosław Bernacki:

   - Analiza stanu światowej literatury na temat prognozowania stężeń zanieczyszczeń powietrza, zaprojektowanie i implementacja metod prognozowania zanieczyszczeń za pomocą sieci neuronowych z uwzględnieniem danych meteorologicznych, przeprowadzenie eksperymentów ewaluacyjnych proponowanych metod z analizą wpływu danych meteorologicznych na jakość prognoz oraz przygotowanie tekstu manuskryptu wraz z opracowaniem elementów wspomagających (tabele, rysunki).

2. Izabela Sówka:

   - Udział w projektowaniu metody prognozowania stężeń zanieczyszczeń powietrza, nadzór naukowy, redakcja tekstu – analiza oraz jego krytyczna recenzja.

Praca A7 – J. Bernacki, I. Sówka. Forecasting Particulate Matter Concentrations with Deep Neural Networks Using Meteorological Data. Zeszyty Naukowe SGSP 2026, No. 98(1)

1. Jarosław Bernacki:

   - Analiza stanu światowej literatury na temat prognozowania stężeń zanieczyszczeń powietrza, zaprojektowanie i implementacja metod prognozowania zanieczyszczeń za pomocą sieci neuronowych z uwzględnieniem danych meteorologicznych, przeprowadzenie eksperymentów ewaluacyjnych proponowanych metod z analizą wpływu danych meteorologicznych na jakość prognoz oraz przygotowanie tekstu manuskryptu wraz z opracowaniem elementów wspomagających (tabele, rysunki).

2. Izabela Sówka:

   - Udział w projektowaniu metody prognozowania stężeń zanieczyszczeń powietrza, nadzór naukowy, redakcja tekstu – analiza oraz jego krytyczna recenzja.

A.11 Oświadczenie autora – Załącznik 11

Oświadczam, że niniejsza rozprawa (dyscyplina: *inżynieria środowiska, górnictwo i energetyka*) prezentuje oryginalne wyniki, które nie zostały przedstawione jako całość lub część mojej wcześniejszej rozprawy doktorskiej pt. „Algorytmy identyfikacji sensorów obrazujących" w dyscyplinie *informatyka techniczna i telekomunikacja* (promotor: prof. dr hab. inż. Rafał Scherer; obrona pracy w dniu 19.10.2023 r., decyzja Rady Wydziału Inżynierii Mechanicznej i Informatyki Politechniki Częstochowskiej z dnia 26.10.2023 r. w sprawie nadania stopnia doktora w dyscyplinie: *informatyka techniczna i telekomunikacja*).

Dodatek B

Poprzednia praca doktorska

W niniejszym rozdziale przedstawiona i krótko scharakteryzowana zostanie wcześniejsza praca doktorska autora, pt. „Algorytmy identyfikacji sensorów obrazujących”, obroniona w dyscyplinie informatyka techniczna i telekomunikacja (ITiT). Przedstawione zostaną główne założenia badawcze poprzedniej pracy, jak również publikacje, które powstały w ramach prowadzonych badań.

B.1 Kryminalistyka cyfrowa

Kryminalistyka cyfrowa (ang. *digital forensics*) to dziedzina zajmująca się analizą danych pochodzących z urządzeń cyfrowych. Informacje te mogą stanowić materiał dowodowy, wykorzystywany między innymi w postępowaniach sądowych. Do urządzeń, z których takie dane mogą mogą pochodzić, należą m.in. komputery (w tym serwery), telefony komórkowe oraz inne urządzenia mobilne, takie jak smartfony czy tablety, a także sprzęt wyposażony w sensory obrazujące, np. aparaty cyfrowe, skanery czy drukarki. Sensor obrazujący jest elementem elektronicznym zawierającym materiały światłoczułe, umożliwiające rejestrowanie obrazu oraz jego zapis w postaci cyfrowej. Tego typu sensory stosowane są m.in. w aparatach cyfrowych, skanerach płaskich oraz w urządzeniach mobilnych posiadających wbudowane aparaty fotograficzne, takich jak smartfony i tablety.

Identyfikacja aparatów cyfrowych na podstawie zdjęć (ang. *digital camera identification*) od wielu lat stanowi istotny kierunek badań. Problem ten rozpatrywany jest w dwóch głównych ujęciach: identyfikacji indywidualnego źródła (ang. *individual source camera identification* – ISCI) oraz identyfikacji modelu aparatu (ang. *source camera model identification* – SCMI). W podejściu ISCI rozróżnia się poszczególne egzemplarze urządzeń, traktując każdy aparat jako odrębne źródło obrazu. Przykładowo, posiadając kilka egzemplarzy modeli aparatów cyfrowych Nikon D810 i Sony A7S, każdy aparat jest identyfikowany indywidualnie, np. Nikon D810 (1), Nikon D810 (2), ... Sony A7S (1), Sony A7S (2), Sony A7S (3), ... Natomiast w metodologii SCMI celem jest jedynie określenie modelu aparatu, z którego pochodzi zdjęcie, bez rozróżniania poszczególnych egzemplarzy tego samego typu. W literaturze naukowej większą uwagę poświęca się metodom opartym na podejściu ISCI.

Rozprawa poruszała temat identyfikacji sensorów obrazujących w oparciu o analizę materiałów, które z nich pochodzą. Rozpatrywanymi urządzeniami były aparaty cyfrowe (lustrzanki, bezlusterkowce, aparaty kompaktowe, a także urządzenia mobilne, takie jak smartfony lub tablety) oraz skanery płaskie. Analiza dotyczyła przede wszystkim identyfikacji w aspekcie ISCI na podstawie naturalnych zdjęć, przedstawiających rzeczywiste sceny (tj. pochodzących „wprost" z aparatu cyfrowego, a więc niepoddawanych zewnętrznym modyfikacjom w oprogramowaniu służącym do obróbki zdjęć, typu Adobe Lightroom/Photoshop itd.).

Analiza literatury wskazała, że główny nacisk w badaniach kładziony jest na uzyskanie jak najwyższej skuteczności identyfikacji sensorów obrazujących. Stosunkowo niewiele prac poświęca jednak uwagę szybkości działania proponowanych algorytmów, koncentrując się niemal wyłącznie na dokładności klasyfikacji. Stanowiło to motywację do podjęcia badań w tym kierunku. **Tezę pracy** sformułowano następująco:

Istnieją algorytmy identyfikacji sensorów obrazujących w aspekcie ISCI, które działają szybciej niż algorytmy istniejące w literaturze przy jednoczesnym zachowaniu porównywalnej skuteczności klasyfikacji.

W związku z tym, **celem rozprawy** było opracowanie algorytmów identyfikacji sensorów obrazujących w ujęciu ISCI, które będą działały szybciej niż rozwiązania opisane w literaturze, przy jednoczesnym zachowaniu porównywalnej skuteczności identyfikacji.

B.2 Publikacje

Rozprawa była oparta na wynikach badań opublikowanych w dziewięciu publikacjach[1] o zasięgu międzynarodowym, indeksowanych w bazie Web of Science, których dane bibliograficzne zostały przedstawione w Tab. B.1.

[1]Na podstawie art. 267 ust. 3 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce (Dz. U. Z 2020 r. poz. 85, 374, 695, 875 i 1086 oraz z 2021 r. poz. 159)

Tabela B.1. Publikacje wchodzące w zakres rozprawy z ITiT. W tabeli podano punktację wg ówczesnej listy MEiN; wartość wskaźnika Impact Factor (IF) określonych czasopism z roku publikacji oraz liczbę cytowań (stan na dzień 1.07.2026 r. wg Google Scholar)

L.p.	Pełny opis bibliograficzny	Pkt.	IF	Cyt.
A1	J. Bernacki (2020): A survey on digital camera identification methods; Forensic Science International: Digital Investigation (Elsevier). 34: 300983	20[2]	2,2	50
B1	J. Bernacki, R. Scherer (2023): IMAGINE Dataset: Digital Camera Identification Image Benchmarking Dataset. Proceedings of International Conference on Security and Cryptography (Springer, SciTePress). 799-804	70	–	10
C1	J. Bernacki, R. Scherer (2021): Fast Imaging Sensor Identification; Proceedings of International Conference on Computational Collective Intelligence (Springer, Lecture Notes in Computer Science). 572-584	70	–	0
C2	J. Bernacki, R. Scherer (2022): Digital forensics: a fast algorithm for a digital sensor identification; Journal of Information and Telecommunication (Taylor & Francis). 6(4): 399-419	20	2,7	5
C3	J. Bernacki (2022): Digital camera identification by fingerprint's compact representation; Multimedia Tools and Applications (Springer). 81(15): 21641-21674	70	3,6	15
C4	J. Bernacki (2020): Digital camera identification based on analysis of optical defects; Multimedia Tools and Applications (Springer). 79(3-4): 2945-2963	70	2,8	9
C5	J. Bernacki, K. Costa, R. Scherer (2022): Individual Source Camera Identification with Convolutional Neural Networks; Proceedings of Asian Conference on Intelligent Information and Database Systems (Springer, Communications in Computer and Information Science), 45-55	70	–	4
D1	J. Bernacki (2021): On robustness of camera identification algorithms; Multimedia Tools and Applications (Springer). 80(1): 921-942	70	2,6	10
D2	J. Bernacki (2021): Robustness of digital camera identification with convolutional neural networks; Multimedia Tools and Applications (Springer). 80(19): 29657-29673	70	2,6	16
	Σ	**530**	**16,4**	**119**

W pracy **A1** będącej artykułem przeglądowym scharakteryzowano ówczesny stan badań na temat identyfikacji sensorów obrazujących. W pracy **B1** zaprezentowano zbiór zdjęć IMAGI-

[2]W trakcie recenzji manuskryptu czasopismo doznało zmiany liczby punktów ze 100 na 20.

NE[3], nadający się do testowania algorytmów identyfikacji zarówno w aspekcie ISCI, jak i SCMI. W publikacjach **C1–C4** przedstawiono kolejne algorytmy identyfikacji sensorów obrazujących, nazwane MSE-DSI (identyfikacja statystyczna), CompaRe (macierzowe przekształcenia cyfrowego odcisku palca), Vignetting-CT (identyfikacja na podstawie wady optycznej winietowania obiektywu) oraz Distortion-CT (identyfikacja na podstawie wady dystorsji obiektywu). Dodatkowo, artykuły **D1** oraz **D2** zawierały analizę odporności algorytmów identyfikacji.

Jako wkład rozprawy można było zatem wskazać:

(1) Opracowanie zbioru zdjęć IMAGINE, zawierającego reprezentatywną liczbę zdjęć pochodzącą z szeregu nowoczesnych urządzeń obrazujących; zbiór został zamieszczony na stronie internetowej do łatwego pobierania;

(2) Opracowanie algorytmów MSE-DSI, CompaRe, Vignetting-CT oraz Distortion-CT, realizujących zadanie identyfikacji sensorów obrazujących na podstawie zdjęć. Algorytmy wykazały się przede wszystkim dużo krótszym czasem przetwarzania zdjęć, w odniesieniu do algorytmów z literatury;

(3) Zbadanie odporności algorytmów oraz konwolucyjnych sieci neuronowych w ramach identyfikacji na podstawie zdjęć poddanych różnorodnym operacjom zaszumiania;

(4) Opracowanie metody przeciwdziałania identyfikacji w rozumieniu algorytmu Lukás [132], wykorzystującej próbkowanie Lanczosa.

Podsumowując, rozległa weryfikacja eksperymentalna pozwoliła dowieść założonej tezy pracy, a cel rozprawy został zrealizowany.

B.3 Dodatkowe informacje

Rozprawa została przygotowana w ramach kształcenia w Szkole Doktorskiej Politechniki Częstochowskiej, które trwało od 1 października 2019 roku do 23 marca 2023 roku (złożenie rozprawy) pod kierunkiem prof. dra hab. inż. Rafała Scherera[4]. Obrona pracy miała miejsce 19 października 2023 roku na Wydziale Inżynierii Mechanicznej i Informatyki, natomiast decyzja Rady Wydziału o nadaniu stopnia doktora w dyscyplinie informatyka techniczna i telekomunikacja oraz wyróżnieniu rozprawy została podjęta 26 października 2023 roku.

[3] Strona zbioru dostępna jest pod adresem: https://ksi.pcz.pl/imagine/

[4] https://ludzie.nauka.gov.pl/ln/profiles/rafa%C5%82.scherer.Sl2Ep9DhG22

Ponadto, autor zgłosił rozprawę do ogólnopolskiego konkursu na najlepsze prace inżynierskie, magisterskie oraz doktorskie organizowanego przez firmę TRUMPF Huettinger (przy współpracy z Politechniką Warszawską). Praca została nagrodzona trzecim miejscem spośród kilkudziesięciu prac doktorskich zgłoszonych z uczelni z całej Polski. Uroczysta gala, na której ogłoszono wyniki (i wręczono nagrody) odbyła się w Małej Auli Politechniki Warszawskiej 13 czerwca 2024 roku[5].

Po obronie, autor przez dwa lata podejmował zatrudnienie na stanowisku adiunkta badawczo-dydaktycznego na Wydziale Inżynierii Mechanicznej i Informatyki (później, Wydziale Informatyki i Sztucznej Inteligencji) Politechniki Częstochowskiej; aktualnie autor zatrudniony jest na stanowisku adiunkta badawczo-dydaktycznego na Wydziale Informatyki i Telekomunikacji Politechniki Wrocławskiej (i nadal aktywnie współpracuje z zespołem badawczym z Częstochowy).

[5] https://www.trumpf.com/filestorage/TRUMPF_PL/Konkurs_2026/LAUREACI_I_FINALISCI_KONKURSU_ZA_ROK_2023.pdf

Spis rysunków

Spis tabel

Bibliografia

[1] Samsuri Abdullah, Marzuki Ismail, Ali Najah Ahmed, and Ahmad Makmom Abdullah. Forecasting particulate matter concentration using linear and non-linear approaches for air quality decision support. *Atmosphere*, 10(11):667, 2019.

[2] Sadia Afrin, Mohammad Maksimul Islam, Tanvir Ahmed, et al. A meteorology based particulate matter prediction model for megacity dhaka. *Aerosol and Air Quality Research*, 21(4):200371, 2021.

[3] Mohammad Hossein Ahmadi, Hamidreza Jashnani, Kwok-Wing Chau, Ravinder Kumar, and Marc A Rosen. Carbon dioxide emissions prediction of five middle eastern countries using artificial neural networks. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 45(3):9513–9525, 2023.

[4] Ali Najah Ahmed, Lee Woen Ean, Ming Fai Chow, Marlinda Abdul Malek, et al. Forecasting of fine particulate matter based on lstm and optimization algorithm. *Journal of Cleaner Production*, 427:139233, 2023.

[5] Aishah Al Yammahi and Zeyar Aung. Forecasting the concentration of no2 using statistical and machine learning methods: A case study in the uae. *Heliyon*, 9(2), 2023.

[6] Erdinç Aladağ. Forecasting of particulate matter with a hybrid arima model based on wavelet transformation and seasonal adjustment. *Urban Climate*, 39:100930, 2021.

[7] Ioannis D. Apostolopoulos, George Fouskas, and Spyros N. Pandis. Field calibration of a low-cost air quality monitoring device in an urban background site using machine learning models. *Atmosphere*, 14(2), 2023.

[8] José L Aznarte. Probabilistic forecasting for extreme no2 pollution episodes. *Environmental Pollution*, 229:321–328, 2017.

[9] Artur Badyda, Anna Gayer, Piotr Oskar Czechowski, Grzegorz Majewski, and Piotr Dąbrowiecki. Pulmonary function and incidence of selected respiratory diseases depending on the exposure to ambient pm10. *International journal of molecular sciences*, 17(11):1954, 2016.

[10] Aleksandra Banasiewicz, Forougholsadat Moosavi, Michalina Kotyla, Paweł Śliwiński, Pavlo Krot, Jacek Wodecki, and Radosław Zimroz. Forecasting of nox emissions of diesel lhd vehicles in underground mines—an ann-based regression approach. *Applied Sciences*, 13(17):9965, 2023.

[11] Dor Bank, Noam Koenigstein, and Raja Giryes. Autoencoders. *Machine learning for data science handbook: data mining and knowledge discovery handbook*, pages 353–374, 2023.

[12] Maximilian Beck, Korbinian Pöppel, Markus Spanring, Andreas Auer, Oleksandra Prudnikova, Michael Kopp, Günter Klambauer, Johannes Brandstetter, and Sepp Hochreiter. xlstm: Extended long short-term memory. *Advances in Neural Information Processing Systems*, 37:107547–107603, 2024.

[13] Abdellatif Bekkar, Badr Hssina, Samira Douzi, and Khadija Douzi. Air-pollution prediction in smart city, deep learning approach. *Journal of big Data*, 8:1–21, 2021.

[14] Jarosław Bernacki. Modelowanie i prognozowanie pyłu zawieszonego z wykorzystaniem hybrydowych autoenkoderów i danych meteorologicznych. *Annales Universitatis Mariae Curie-Sklodowska, sectio B–Geographia, Geologia, Mineralogia et Petrographia*, 80:265–283.

[15] Jarosław Bernacki. Forecasting the air pollution concentration with neural networks. *Urban Climate*, 59:102262, 2025.

[16] Jarosław Bernacki. Forecasting the concentration of the components of the particulate matter in poland using neural networks. *Environmental Science and Pollution Research*, pages 1–34, 2025.

[17] Jarosław Bernacki. A hybrid transformer and symbolic regression model for weather-dependent particulate matter forecasting in air quality management. *Theoretical and Applied Climatology*, 156(12):662, 2025.

[18] Jaroslaw Bernacki and Rafał Scherer. A comprehensive review of data-driven techniques for air pollution concentration forecasting. *Sensors*, 25(19):6044, 2025.

[19] Jarosław Bernacki and Izabela Sówka. Forecasting particulate matter concentrations with deep neural networks using meteorological data. Praca zaakceptowana do czasopisma Scientific Reports of Fire University, 2026.

[20] Jarosław Bernacki and Izabela Sówka. Forecasting particulate matter with neural networks: Accuracy and explainability using meteorological data. Praca zaakceptowana na 25th International Conference on Artificial Intelligence and Soft Computing, 2026.

[21] Uzair Aslam Bhatti, Yuhuan Yan, Mingquan Zhou, Sajid Ali, Aamir Hussain, Huo Qingsong, Zhaoyuan Yu, and Linwang Yuan. Time series analysis and forecasting of air pollution particulate matter (pm 2.5): an sarima and factor analysis approach. *Ieee Access*, 9:41019–41031, 2021.

[22] Ewa Błaszczyk, Wioletta Rogula-Kozłowska, Krzysztof Klejnowski, Piotr Kubiesa, Izabela Fulara, and Danuta Mielżyńska-Švach. Indoor air quality in urban and rural kindergartens: short-term studies in silesia, poland. *Air Quality, Atmosphere & Health*, 10(10):1207–1220, 2017.

[23] Peter J Brockwell and Richard A Davis. *Time series: theory and methods*. Springer science & business media, 2009.

[24] Krzysztof Brzozowski, Łukasz Drąg, and Lucyna Brzozowska. The fogees system for forecasting particulate matter concentrations in urban areas. *Environmental Modelling & Software*, 183:106205, 2025.

[25] Sangwon Chae, Joonhyeok Shin, Sungjun Kwon, Sangmok Lee, Sungwon Kang, and Donghyun Lee. Pm10 and pm2. 5 real-time prediction models using an interpolated convolutional neural network. *Scientific Reports*, 11(1):11952, 2021.

[26] Yue-Shan Chang, Satheesh Abimannan, Hsin-Ta Chiao, Chi-Yeh Lin, and Yo-Ping Huang. An ensemble learning based hybrid model and framework for air pollution forecasting. *Environmental Science and Pollution Research*, 27:38155–38168, 2020.

[27] Yue-Shan Chang, Hsin-Ta Chiao, Satheesh Abimannan, Yo-Ping Huang, Yi-Ting Tsai, and Kuan-Ming Lin. An lstm-based aggregated model for air pollution forecasting. *Atmospheric Pollution Research*, 11(8):1451–1463, 2020.

[28] Fu Chen, Wanyue Zhang, Manar Fawzi Bani Mfarrej, Muhammad Hamzah Saleem, Khalid Ali Khan, Jing Ma, António Raposo, and Heesup Han. Breathing in danger: Under-

standing the multifaceted impact of air pollution on health impacts. *Ecotoxicology and Environmental Safety*, 280:116532, 2024.

[29] Ling Chen, Jiahui Xu, Binqing Wu, and Jianlong Huang. Group-aware graph neural network for nationwide city air quality forecasting. *ACM Trans. Knowl. Discov. Data*, 18(3), December 2023.

[30] Shuixia Chen, Jian qiang Wang, and Hong yu Zhang. A hybrid pso-svm model based on clustering algorithm for short-term atmospheric pollutant concentration forecasting. *Technological Forecasting and Social Change*, 146:41–54, 2019.

[31] Wei Chen, Hongzhao Tang, and Haimeng Zhao. Urban air quality evaluations under two versions of the national ambient air quality standards of china. *Atmospheric Pollution Research*, 7(1):49–57, 2016.

[32] Wan-Li Cheng, Yu-Song Chen, Junfeng Zhang, TJ Lyons, Joy-Lynn Pai, and Shiang-Hung Chang. Comparison of the revised air quality index with the psi and aqi indices. *Science of the Total Environment*, 382(2-3):191–198, 2007.

[33] Davide Chicco, Matthijs J Warrens, and Giuseppe Jurman. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *Peerj computer science*, 7:e623, 2021.

[34] Ji Hyeon Cho and Jin Woo Moon. Integrated artificial neural network prediction model of indoor environmental quality in a school building. *Journal of Cleaner Production*, 344:131083, 2022.

[35] Robert Cichowicz, Grzegorz Wielgosiński, and Wojciech Fetter. Effect of wind speed on the level of particulate matter pm10 concentration in atmospheric air during winter season in vicinity of large combustion plant. *Journal of Atmospheric Chemistry*, 77(1):35–48, 2020.

[36] Yagmur Gizem Cinar, Hamid Mirisaee, Parantapa Goswami, Eric Gaussier, Ali Aït-Bachir, and Vadim Strijov. Position-based content attention for time series forecasting with sequence-to-sequence rnns. In *Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14–18, 2017, Proceedings, Part V 24*, pages 533–544. Springer, 2017.

[37] Bartosz Czernecki, Michał Marosz, and Joanna Jędruszkiewicz. Assessment of machine learning algorithms in short-term forecasting of pm10 and pm2. 5 concentrations in selected polish agglomerations. *Aerosol and Air Quality Research*, 21(7):200586, 2021.

[38] Abdelkader Dairi, Fouzi Harrou, Sofiane Khadraoui, and Ying Sun. Integrated multiple directed attention-based deep learning for improved air pollution forecasting. *IEEE Transactions on Instrumentation and Measurement*, 70:1–15, 2021.

[39] Saverio De Vito, Gerardo D'Elia, Sergio Ferlito, Girolamo Di Francia, Miloš D. Davidović, Duška Kleut, Danka Stojanović, and Milena Jovaševic-Stojanović. A global multiunit calibration as a method for large-scale iot particulate matter monitoring systems deployments. *IEEE Transactions on Instrumentation and Measurement*, 73:1–16, 2024.

[40] Rahul Dey and Fathi M Salem. Gate-variants of gated recurrent unit (gru) neural networks. In *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*, pages 1597–1600. IEEE, 2017.

[41] Luis A Díaz-Robles, Juan C Ortega, Joshua S Fu, Gregory D Reed, Judith C Chow, John G Watson, and Juan A Moncada-Herrera. A hybrid arima and artificial neural networks model to forecast particulate matter in urban areas: The case of temuco, chile. *Atmospheric Environment*, 42(35):8331–8340, 2008.

[42] Robert DiPietro and Gregory D Hager. Deep learning: Rnns and lstm. In *Handbook of medical image computing and computer assisted intervention*, pages 503–519. Elsevier, 2020.

[43] Mahmut Dirik. Prediction of nox emissions from gas turbines of a combined cycle power plant using an anfis model optimized by ga. *Fuel*, 321:124037, 2022.

[44] Marius Dobrea, Andreea Bădicu, Marina Barbu, Oana Subea, Mihaela Bălănescu, Geroge Suciu, Andrei Bîrdici, Oana Orza, and Ciprian Dobre. Machine learning algorithms for air pollutants forecasting. In *2020 IEEE 26th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, pages 109–113. IEEE, 2020.

[45] Carl Doersch. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*, 2016.

[46] Jie Dong, Yaoli Zhang, and Jiang Hu. Short-term air quality prediction based on emd-transformer-bilstm. *Scientific Reports*, 14(1):20513, 2024.

[47] Philip J Drew and John RT Monson. Artificial neural networks. *Surgery*, 127(1):3–11, 2000.

[48] Pei Du, Jianzhou Wang, Wendong Yang, and Tong Niu. A novel hybrid fine particulate matter (pm2. 5) forecasting and its further application system: Case studies in china. *Journal of Forecasting*, 41(1):64–85, 2022.

[49] Shengdong Du, Tianrui Li, Yan Yang, and Shi-Jinn Horng. Deep air quality forecasting using hybrid deep learning framework. *IEEE Transactions on Knowledge and Data Engineering*, 33(6):2412–2424, 2019.

[50] Radhika Dua Dua, Divyam Madaan Madaan, Prerana Mukherjee Mukherjee, and Brejesh Lall Lall. Real time attention based bidirectional long short-term memory networks for air pollution forecasting. In *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 151–158, 2019.

[51] Jiahui Duan, Yaping Gong, Jun Luo, and Zhiyao Zhao. Air-quality prediction based on the arima-cnn-lstm combination model optimized by dung beetle optimizer. *Scientific Reports*, 13(1):12127, 2023.

[52] Rita M Durao, Manuel T Mendes, and M Joao Pereira. Forecasting o3 levels in industrial area surroundings up to 24 h in advance, combining classification trees and mlp models. *Atmospheric Pollution Research*, 7(6):961–970, 2016.

[53] Sbai Salah Eddine, Lalla Btissam Drissi, Nezha Mejjad, Jamal Mabrouki, and Aleksey A Romanov. Machine learning models application for spatiotemporal patterns of particulate matter prediction and forecasting over morocco in north of africa. *Atmospheric Pollution Research*, 15(9):102239, 2024.

[54] Khalid Elbaz, Wafaa Mohamed Shaban, Annan Zhou, and Shui-Long Shen. Real time image-based air quality forecasts using a 3d-cnn approach with an attention mechanism. *Chemosphere*, 333:138867, 2023.

[55] Philippe Esling and Carlos Agon. Time-series data mining. *ACM Computing Surveys (CSUR)*, 45(1):1–34, 2012.

[56] Junxiang Fan, Qi Li, Junxiong Hou, Xiao Feng, Hamed Karimian, and Shaofu Lin. A spatiotemporal prediction framework for air pollution based on deep rnn. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4:15–22, 2017.

[57] Gema Fernández-Avilés, Raffaele Mattera, and Germana Scepi. Factor-augmented autoregressive neural network to forecast nox in the city of madrid. *Socio-Economic Planning Sciences*, page 101958, 2024.

[58] Mariusz Filak. Predykcja stężeń zanieczyszczeń powietrza w perceptronowych modelach regresyjnych. Rozprawa doktorska, Politechnika Częstochowska, Częstochowa, 2022.

[59] Crescenzio Gallo, Francesco Conto, and Mariantonietta Fiore. A neural network model for forecasting co2 emission. *AGRIS on-line Papers in Economics and Informatics*, 6(2):31–36, 2014.

[60] Wenyuan Gao, Tongjue Xiao, Lin Zou, Huan Li, and Shengbo Gu. Analysis and prediction of atmospheric environmental quality based on the autoregressive integrated moving average model (arima model) in hunan province, china. *Sustainability*, 16(19):8471, 2024.

[61] María Vega García and José L Aznarte. Shapley additive explanations for no2 forecasting. *Ecological Informatics*, 56:101039, 2020.

[62] F WERNER Gerstengarbe and PC Werner. A method for the statistical definition of extreme-value regions and their application to meteorological time series. *Zeitschrift fuer Meteorologie;(German Democratic Republic)*, 39(4), 1989.

[63] Aysenur Gilik, Arif Selcuk Ogrenci, and Atilla Ozmen. Air quality prediction using cnn+ lstm-based hybrid deep learning architecture. *Environmental science and pollution research*, pages 1–19, 2022.

[64] Priyanjali Gogikar, Bhishma Tyagi, and AK Gorai. Seasonal prediction of particulate matter over the steel city of india using neural network models. *Modeling Earth Systems and Environment*, 5:227–243, 2019.

[65] Karolina Gora and Mateusz Rzeszutek. Comparison of selected ensemble supervised learning algorithms used for meteorological normalisation of particulate matter (pm 10). *Sustainability*, 17(12):1–17, 2025.

[66] Tomasz Gorzelnik, Marek Bogacki, and Robert Oleniacz. Identification of factors influencing episodes of high pm10 concentrations in the air in krakow (poland) using random forest method. *Sustainability*, 16(20):9015, 2024.

[67] Dejan Gradišar, Boštjan Grašič, Marija Zlata Božnar, Primož Mlakar, and Juš Kocijan. Improving of local ozone forecasting by integrated models. *Environmental Science and Pollution Research*, 23:18439–18450, 2016.

[68] Alex Graves. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, pages 37–45, 2012.

[69] Yuanlin Gu, Baihua Li, and Qinggang Meng. Hybrid interpretable predictive machine learning model for air pollution prediction. *Neurocomputing*, 468:123–136, 2022.

[70] Qingchun Guo, Zhenfang He, Shanshan Li, Xinzhou Li, Jingjing Meng, Zhanfang Hou, Jiazhen Liu, and Yongjin Chen. Air pollution forecasting using artificial and wavelet neural networks with meteorological conditions. *Aerosol and Air Quality Research*, 20(6):1429–1439, 2020.

[71] Qingchun Guo, Zhenfang He, and Zhaosheng Wang. Predicting of daily pm2. 5 concentration employing wavelet artificial neural networks based on meteorological elements in shanghai, china. *Toxics*, 11(1):51, 2023.

[72] Qingchun Guo, Zhenfang He, and Zhaosheng Wang. Prediction of hourly pm2.5 and pm10 concentrations in chongqing city in china based on artificial neural network. *Aerosol and Air Quality Research*, 23(6):220448, 2023.

[73] Xifeng Guo, Xinwang Liu, En Zhu, and Jianping Yin. Deep clustering with convolutional autoencoders. In *International conference on neural information processing*, pages 373–382. Springer, 2017.

[74] Główny Inspektorat Ochrony Środowiska (GIOŚ). https://www.gov.pl/web/gios, 2025. Dostęp: 23 grudnia 2025.

[75] Stephen Haben, Marcus Voss, and William Holderbaum. Time series forecasting: Core concepts and definitions. In *Core Concepts and Methods in Load Forecasting: With Applications in Distribution Networks*, pages 55–66. Springer, 2023.

[76] Ghassan B. Hamra, Neela Guha, Aaron Cohen, Francine Laden, Ole Raaschou-Nielsen, Jonathan M. Samet, Paolo Vineis, Francesco Forastiere, Paulo Saldiva, Takashi Yorifuji, and Dana Loomis. Outdoor particulate matter exposure and lung cancer: A systematic review and meta-analysis. *Environmental Health Perspectives*, 122(9):906–911, 2014.

[77] Jindong Han, Hao Liu, Haoyi Xiong, and Jing Yang. Semi-supervised air quality forecasting via self-supervised hierarchical graph neural network. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):5230–5243, 2023.

[78] Zhongyang Han, Jun Zhao, Henry Leung, King Fai Ma, and Wei Wang. A review of deep learning models for time series prediction. *IEEE Sensors Journal*, 21(6):7833–7848, 2019.

[79] KS Harishkumar, KM Yogesh, Ibrahim Gad, et al. Forecasting air pollution particulate matter (pm2. 5) using machine learning regression models. *Procedia Computer Science*, 171:2057–2066, 2020.

[80] Zhenfang He, Qingchun Guo, Zhaosheng Wang, and Xinzhou Li. Prediction of monthly pm2. 5 concentration in liaocheng in china employing artificial neural network. *Atmosphere*, 13(8):1221, 2022.

[81] Azim Heydari, Meysam Majidi Nezhad, Davide Astiaso Garcia, Farshid Keynia, and Livio De Santoli. Air pollution forecasting application based on deep learning model and optimization algorithm. *Clean Technologies and Environmental Policy*, pages 1–15, 2022.

[82] Sebastian H. M. Hickman, Paul T. Griffiths, Peer J. Nowack, and Alexander T. Archibald. Short-term forecasting of ozone air pollution across europe with transformers. *Environmental Data Science*, 2:e43, 2023.

[83] Szymon Hoffman and Rafał Jasiński. The use of multilayer perceptrons to model pm2. 5 concentrations at air monitoring stations in poland. *Atmosphere*, 14(1):96, 2023.

[84] Ping Wang Yuzhe Qin Hong Zhang, Sheng Zhang and Huifeng Wang. Forecasting of particulate matter time series using wavelet analysis and wavelet-arma/arima model in taiyuan, china. *Journal of the Air & Waste Management Association*, 67(7):776–788, 2017. PMID: 28278031.

[85] Seyed Mohsen Hosseini, Amirali Saifoddin, Reza Shirmohammadi, and Alireza Aslani. Forecasting of co2 emissions in iran based on time series and regression analysis. *Energy Reports*, 5:619–631, 2019.

[86] Yuan Huang, Yuxing Xiang, Ruixiao Zhao, and Zhe Cheng. Air quality prediction using improved pso-bp neural network. *Ieee Access*, 8:99346–99353, 2020.

[87] Ditsuhi Iskandaryan, Francisco Ramos, and Sergio Trilles. Graph neural network for air quality prediction: A case study in madrid. *IEEE Access*, 11:2729–2742, 2023.

[88] Sebastian Iwaszenko, Adam Smolinski, Marcin Grzanka, and Tomasz Skowronek. Airborne particulate matter measurement and prediction with machine learning techniques. *Scientific Reports*, 14(1):18999, 2024.

[89] Agata Jaroń, Anna Borucka, and Maciej Ślusarczyk. Long-term trends in air pollution in poland on selected examples—a spatial and temporal analysis of pm10 concentrations. *Applied Sciences*, 15(23):12379, 2025.

[90] Agata Jaroń, Anna Borucka, Paulina Deliś, and Aleksandra Sekrecka. An assessment of the possibility of using unmanned aerial vehicles to identify and map air pollution from infrastructure emissions. *Energies*, 17(3), 2024.

[91] Eun-Jung Jo, Woo-Seop Lee, Hyun-Young Jo, Chang-Hoon Kim, Jung-Seop Eom, Jeong-Ha Mok, Mi-Hyun Kim, Kwangha Lee, Ki-Uk Kim, Min-Ki Lee, and Hye-Kyung Park. Effects of particulate matter on respiratory disease and the impact of meteorological factors in busan, korea. *Respiratory Medicine*, 124:79–87, 2017.

[92] Thomas Johnson and Kieran Woodward. Enviro-iot: calibrating low-cost environmental sensors in urban settings. *arXiv preprint arXiv:2502.07596*, 2025.

[93] István Juhos, László Makra, and Balázs Tóth. Forecasting of traffic origin no and no2 concentrations by support vector machines and neural networks using principal component analysis. *Simulation Modelling Practice and Theory*, 16(9):1488–1502, 2008.

[94] Federico Karagulian, Maurizio Barbiere, Alexander Kotsev, Laurent Spinelle, Michel Gerboles, Friedrich Lagler, Nathalie Redon, Sabine Crunaire, and Annette Borowiak. Review of the performance of low-cost sensors for air quality monitoring. *Atmosphere*, 10(9):506, 2019.

[95] Gagandeep Kaur and Harmanpreet Kaur. Ensemble based j48 and random forest based c6h6 air pollution detection. *International Journal of Scientific Research in Computer Science and Engineering*, 6(2):41–50, 2018.

[96] Huam Ming Ken, Mehran Behjati, Ahmad Sahban Rafsanjani, Saad Aslam, Yap Kian Meng, Anwar P. P. Abdul Majeed, and Yufan Zheng. *Advancing Air Quality Monitoring: TinyML-Based Real-Time Ozone Prediction with Cost-Effective Edge Devices*, page 502–512. Springer Nature Singapore, 2025.

[97] Nikhil Ketkar and Jojo Moolayil. Convolutional neural networks. In *Deep learning with Python: learn best practices of deep learning models with PyTorch*, pages 197–242. Springer, 2021.

[98] Joanna Kobza, Mariusz Geremek, and Lechosław Dul. Characteristics of air quality and sources affecting high levels of pm10 and pm2. 5 in poland, upper silesia urban area. *Environmental monitoring and assessment*, 190(9):515, 2018.

[99] Wojciech Konicki, Milena Bojanowska, Anna Eliza Wolnowska, Karolina Nadolska, Urszula Gabriel-Półrolniczak, and Anna Kokorniak. Forecasting pm10 concentration using ann narx from road transport in szczecin, poland. *Scientific Papers of Silesian University of Technology. Organization & Management/Zeszyty Naukowe Politechniki Slaskiej. Seria Organizacji i Zarzadzanie*, (230), 2025.

[100] Georgios N Kouziokas. Svm kernel based on particle swarm optimized vector and bayesian optimized svm in atmospheric particulate matter forecasting. *Applied Soft Computing*, 93:106410, 2020.

[101] Piotr A Kowalski, Kasper Sapała, and Wiktor Warchałowski. Pm10 forecasting through applying convolution neural network techniques. *Air Pollution Studies; WIT Press: Southampton, UK*, page 47, 2020.

[102] Ewa Agnieszka Krajny, Leszek Ośródka, and Marek Jan Wojtylak. Application of doppler sodar in short-term forecasting of pm 10 concentration in the air in krakow (poland). *Atmospheric Measurement Techniques*, 17(8):2451–2464, 2024.

[103] Anders Krogh. What are artificial neural networks? *Nature biotechnology*, 26(2):195–197, 2008.

[104] WH Kruskal and WA Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, pages 583–621, 1952.

[105] William H. Kruskal. A Nonparametric test for the Several Sample Problem. *The Annals of Mathematical Statistics*, 23(4):525 – 540, 1952.

[106] Maciej Kryza, Małgorzata Werner, and Jakub Guzikowski. Assimilation of meteorological data in online integrated atmospheric transport model—example of air quality forecasts for poland. In *Air Pollution Modeling and its Application XXVI 36*, pages 273–278. Springer, 2020.

[107] Justyna Kujawska, Monika Kulisz, Piotr Oleszczuk, and Wojciech Cel. Machine learning methods to forecast the concentration of pm10 in lublin, poland. *Energies*, 15(17):6428, 2022.

[108] Gulshan Kumar, D Prasannaa Kumar, Jay Dhariwal, and Seshan Srirangarajan. Reliability assessment of low-cost pm sensors under high humidity and high pm level outdoor conditions. *IEEE Sensors Journal*, 2025.

[109] Ujjwal Kumar and VK Jain. Arima forecasting of ambient air pollutants (o 3, no, no 2 and co). *Stochastic Environmental Research and Risk Assessment*, 24:751–760, 2010.

[110] Gamze Kurnaz and Alparslan Serhat Demir. Prediction of so2 and pm10 air pollutants using a deep learning-based recurrent neural network: Case of industrial city sakarya. *Urban Climate*, 41:101051, 2022.

[111] Marek Lasiewicz, Małgorzata Bogusz, and Magdalena Kośla. Universal system for forecasting changes in pm 10 and pm 2.5 particulate matter air pollution concentration. *International Journal of Environment and Pollution*, 56(1-4):94–108, 2014.

[112] Colin Lea, Michael D Flynn, Rene Vidal, Austin Reiter, and Gregory D Hager. Temporal convolutional networks for action segmentation and detection. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 156–165, 2017.

[113] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[114] Jeong-Young Lee, Wan-Kuen Jo, and Ho-Hwan Chun. Long-term trends in visibility and its relationship with mortality, air-quality index, and meteorological factors in selected areas of korea. *Aerosol and Air Quality Research*, 15(2):673–681, 2015.

[115] Mike Lee, Larry Lin, Chih-Yuan Chen, Yu Tsao, Ting-Hsuan Yao, Min-Han Fei, and Shih-Hau Fang. Forecasting air quality in taiwan by using machine learning. *Scientific reports*, 10(1):4153, 2020.

[116] Pius Lee, Jeffery McQueen, Ivanka Stajner, Jianping Huang, Li Pan, Daniel Tong, Hyuncheol Kim, Youhua Tang, Shobha Kondragunta, Mark Ruminski, et al. Naqfc developmental forecast guidance for fine particulate matter (pm 2.5). *Weather and Forecasting*, 32(1):343–360, 2017.

[117] Sangdeok Lee and MinChung Kim. Application of long-short term memory (lstm) model for forecasting nox emission in pohang area. *arXiv preprint arXiv:2311.15632*, 2023.

[118] Thomas M. T. Lei, Stanley C. W. Ng, and Shirley W. I. Siu. Application of ann, xgboost, and other ml methods to forecast air quality in macau. *Sustainability*, 15(6), 2023.

[119] Anita Urszula Lewandowska and Lucyna Mirosława Falkowska. High concentration episodes of pm10 in the air over the urbanized coastal zone of the baltic sea (gdynia—poland). *Atmospheric Research*, 120:55–67, 2013.

[120] Taoying Li, Miao Hua, and XU Wu. A hybrid cnn-lstm model for forecasting particulate matter (pm2. 5). *Ieee Access*, 8:26933–26940, 2020.

[121] Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng, and Jun Zhou. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 33(12):6999–7019, 2021.

[122] Zeyue Li, Jianzhao Bi, Yang Liu, and Xuefei Hu. Forecasting o3 and no2 concentrations with spatiotemporally continuous coverage in southeastern china using a machine learning approach. *Environment International*, 195:109249, 2025.

[123] Yanjie Liang, Jun Ma, Chuanyang Tang, Nan Ke, and Dong Wang. Hourly forecasting on pm2. 5 concentrations using a deep neural network with meteorology inputs. *Environmental Monitoring and Assessment*, 195(12):1510, 2023.

[124] Yuxuan Liang, Yutong Xia, Songyu Ke, Yiwei Wang, Qingsong Wen, Junbo Zhang, Yu Zheng, and Roger Zimmermann. Airformer: Predicting nationwide air quality in china with transformers. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 14329–14337, 2023.

[125] Tianyang Lin, Yuxin Wang, Xiangyang Liu, and Xipeng Qiu. A survey of transformers. *AI open*, 3:111–132, 2022.

[126] Bingchun Liu, Xiaogang Yu, Jiali Chen, and Qingshan Wang. Air pollution concentration forecasting based on wavelet transform and combined weighting forecasting model. *Atmospheric Pollution Research*, 12(8):101144, 2021.

[127] Bo Liu, Shuo Yan, Jianqiang Li, Guangzhi Qu, Yong Li, Jianlei Lang, and Rentao Gu. An attention-based air quality forecasting method. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 728–733, 2018.

[128] Chih-Chun Liu, Tzu-Chi Lin, Kuang-Yu Yuan, and Pei-Te Chiueh. Spatio-temporal prediction and factor identification of urban air quality using support vector machine. *Urban climate*, 41:101055, 2022.

[129] Duen-Ren Liu, Shin-Jye Lee, Yang Huang, and Chien-Ju Chiu. Air pollution forecasting based on attention-based lstm neural network and ensemble learning. *Expert Systems*, 37(3):e12511, 2020.

[130] Hui Liu, Kairong Jin, and Zhu Duan. Air pm2. 5 concentration multi-step forecasting using a new hybrid modeling method: comparing cases for four cities in china. *Atmospheric Pollution Research*, 10(5):1588–1600, 2019.

[131] Xiaoting Liu, Rohan Jayaratne, Phong Thai, Tara Kuhn, Isak Zing, Bryce Christensen, Riki Lamont, Matthew Dunbabin, Sicong Zhu, Jian Gao, et al. Low-cost sensors as an alternative for long-term air quality monitoring. *Environmental research*, 185:109438, 2020.

[132] Jan Lukás, Jessica J. Fridrich, and Miroslav Goljan. Digital camera identification from sensor pattern noise. *IEEE Trans. Inf. Forensics Secur.*, 1(2):205–214, 2006.

[133] Tao Lyu, Yu Gan, Ru Zhang, Shun Wang, Donghai Li, and Yuqun Zhuo. Development of a real-time nox prediction soft sensor algorithm for power plants based on a hybrid boost integration model. *Energies*, 17(19):4926, 2024.

[134] Jinghui Ma, Zhongqi Yu, Yuanhao Qu, Jianming Xu, and Yu Cao. Application of the xgboost machine learning method in pm2. 5 prediction: A case study of shanghai. *Aerosol and Air Quality Research*, 20(1):128–138, 2020.

[135] Grzegorz Majewski, Wioletta Rogula-Kozłowska, Katarzyna Rozbicka, Patrycja Rogula-Kopiec, Barbara Mathews, and Andrzej Brandyk. Concentration, chemical composition and origin of pm1: Results from the first long-term measurement campaign in Warsaw (Poland). *Aerosol and air quality research*, 18(3):636–654, 2018.

[136] Hugo Mak. *Improved remote sensing algorithms and data assimilation approaches in solving environmental retrieval problems*. PhD thesis, 07 2019.

[137] Hugo Wai Leung Mak and Daisy Chiu Yi Ng. Spatial and socio-classification of traffic pollutant emissions and associated mortality rates in high-density hong kong via improved data analytic approaches. *International Journal of Environmental Research and Public Health*, 18(12), 2021.

[138] Geetha Mani, Joshi Kumar Viswanadhapalli, and Albert Alexander Stonier. Prediction and forecasting of air quality index in chennai using regression and arima time series models. *Journal of Engineering Research*, 10(2):179–194, 2022.

[139] Wenjing Mao, Limin Jiao, Weilin Wang, Jianlong Wang, Xueli Tong, and Suli Zhao. A hybrid integrated deep learning model for predicting various air pollutants. *GIScience & Remote Sensing*, 58(8):1395–1412, 2021.

[140] I Martínez-Silva, Javier Roca-Pardiñas, and C Ordóñez. Forecasting so2 pollution incidents by means of quantile curves based on additive models. *Environmetrics*, 27(3):147–157, 2016.

[141] Patrick E McKight and Julius Najab. Kruskal-wallis test. *The corsini encyclopedia of psychology*, pages 1–1, 2010.

[142] Manuel Méndez, Mercedes G Merayo, and Manuel Núñez. Machine learning algorithms to forecast air quality: a survey. *Artificial Intelligence Review*, 56(9):10031–10066, 2023.

[143] Axel Gedeon Mengara Mengara, Eunyoung Park, Jinho Jang, and Younghwan Yoo. Attention-based distributed deep learning model for air quality forecasting. *Sustainability*, 14(6), 2022.

[144] Meteostat. `https://meteostat.net/en/`, 2025. Dostęp: 23 grudnia 2025.

[145] Dhirendra Mishra and Pramila Goyal. Development of artificial intelligence based no2 forecasting models at taj mahal, agra. *Atmospheric pollution research*, 6(1):99–106, 2015.

[146] Arkadiusz Moskal, Wiktor Jagodowicz, Agata Penconek, and Krzysztof Zaraska. Low-cost sensor system for air purification process evaluation. *Sensors*, 24(6), 2024.

[147] HamidReza Naseri and Vahid Mehrdad. Novel cnn with investigation on accuracy by modifying stride, padding, kernel size and filter numbers. *Multimedia Tools and Applications*, 82(15):23673–23691, 2023.

[148] Paulo SG De Mattos Neto, Paulo Renato A Firmino, Hugo Siqueira, Yara De Souza Tadano, Thiago Antonini Alves, Joao Fausto L De Oliveira, Manoel Henrique Da Nóbrega Marinho, and Francisco Madeiro. Neural-based ensembles for particulate matter forecasting. *IEEE Access*, 9:14470–14490, 2021.

[149] Kai Yang Ng, Duc Thang Ngo, Paul C-P Chao, Ray Hua Horng, and Jia-Min Shieh. High-precision prediction of no x, no 2 and c 6 h 6 by multiple gas sensors using a novel cascaded mlp-lstm model. In *2024 IEEE SENSORS*, pages 1–4. IEEE, 2024.

[150] Jadwiga Nidzgorska-Lencewicz. Application of artificial neural networks in the prediction of pm10 levels in the winter months: a case study in the tricity agglomeration, poland. *Atmosphere*, 9(6):203, 2018.

[151] Jeffrey Ofosu-Adarkwa, Naiming Xie, and Saad Ahmed Javed. Forecasting co2 emissions of china's cement industry using a hybrid verhulst-gm (1, n) model and emissions' technical conversion. *Renewable and Sustainable Energy Reviews*, 130:109945, 2020.

[152] Victor Oliveira Santos, Paulo Alexandre Costa Rocha, John Scott, Jesse Van Griensven Thé, and Bahram Gharabaghi. Spatiotemporal air pollution forecasting in houston-tx: A case study for ozone using deep graph neural networks. *Atmosphere*, 14(2), 2023.

[153] Keiron O'shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[154] Eren Özceylan. Forecasting co2 emission of turkey: swarm intelligence approaches. *International Journal of Global Warming*, 9(3):337–361, 2016.

[155] Mariusz Pagowski, GA Grell, SA McKeen, SE Peckham, and D Devenyi. Three-dimensional variational data assimilation of ozone and fine particulate matter observations: some results using the weather research and forecasting—chemistry model and grid-point statistical interpolation. *Quarterly Journal of the Royal Meteorological Society*, 136(653):2013–2024, 2010.

[156] Unjin Pak, Jun Ma, Unsok Ryu, Kwangchol Ryom, U Juhyok, Kyongsok Pak, and Chanil Pak. Deep learning-based pm2. 5 prediction considering the spatiotemporal correlations: A case study of beijing, china. *Science of the Total Environment*, 699:133561, 2020.

[157] Francisco David Parra-Medina, Manuel Andrés Vélez-Guerrero, and Mauro Callejas-Cuervo. Low-cost solution for air quality monitoring: Unmanned aerial system and data transmission via lora protocol. *Sustainability*, 16(22), 2024.

[158] Izabela Pawlak and Janusz Jarosławski. Forecasting of surface ozone concentration by using artificial neural networks in rural and urban areas in central poland. *Atmosphere*, 10(2):52, 2019.

[159] Państwowy Monitoring Środowiska – Główny Inspektorat Ochrony Środowiska . https://www.gov.pl/web/gios/panstwowy-monitoring-srodowiska, 2025. Dostęp: 23 grudnia 2025.

[160] Walter Hugo Lopez Pinaya, Sandra Vieira, Rafael Garcia-Dias, and Andrea Mechelli. Autoencoders. In *Machine learning*, pages 193–208. Elsevier, 2020.

[161] Sławomir Pochwała, Arkadiusz Gardecki, Piotr Lewandowski, Viola Somogyi, and Stanisław Anweiler. Developing of low-cost air pollution sensor—measurements with the unmanned aerial vehicles in poland. *Sensors*, 20(12), 2020.

[162] Bernard Połednik. Emissions of air pollution in industrial and rural region in poland and health impacts. *Journal of Ecological Engineering*, 23(9):250–258, 2022.

[163] William C Porter, Colette L Heald, Daniel Cooley, and Brook Russell. Investigating the observed sensitivities of air-quality extremes to meteorological drivers via quantile regression. *Atmospheric Chemistry and Physics*, 15(18):10349–10366, 2015.

[164] Kanchan Prasad, Amit Kumar Gorai, and Pramila Goyal. Development of anfis models for air quality forecasting and input optimization for reducing the computational cost and time. *Atmospheric environment*, 128:246–262, 2016.

[165] Yanlin Qi, Qi Li, Hamed Karimian, and Di Liu. A hybrid model for spatiotemporal forecasting of pm2. 5 based on graph convolutional neural network and long short-term memory. *Science of the Total Environment*, 664:1–10, 2019.

[166] Shanshan Qin, Feng Liu, Jianzhou Wang, and Beibei Sun. Analysis and forecasting of the particulate matter (pm) concentration levels over four major cities of china using hybrid models. *Atmospheric Environment*, 98:665–675, 2014.

[167] Enislay Ramentol, Stefanie Grimm, Moritz Stinzendörfer, and Andreas Wagner. Short-term air pollution forecasting using embeddings in neural networks. *Atmosphere*, 14(2):298, 2023.

[168] Gokulan Ravindiran, Gasim Hayder, Karthick Kanagarathinam, Avinash Alagumalai, and Christian Sonne. Air quality prediction by machine learning models: A predictive study on the indian coastal city of visakhapatnam. *Chemosphere*, 338:139518, 2023.

[169] Wioletta Rogula-Kozłowska, Krzysztof Klejnowski, Patrycja Rogula-Kopiec, Leszek Ośródka, Ewa Krajny, Barbara Błaszczak, and Barbara Mathews. Spatial and seasonal variability of the mass concentration and chemical composition of pm2. 5 in poland. *Air Quality, Atmosphere & Health*, 7(1):41–58, 2014.

[170] Raul Rojas. The backpropagation algorithm. In *Neural networks: a systematic introduction*, pages 149–182. Springer, 1996.

[171] Seyed Mostafa Safdarnejad, Jake F Tuttle, and Kody M Powell. Dynamic modeling and optimization of a coal-fired utility boiler to forecast and minimize nox and co emissions simultaneously. *Computers & Chemical Engineering*, 124:62–79, 2019.

[172] A Suárez Sánchez, PJ García Nieto, P Riesgo Fernández, JJ del Coz Díaz, and Francisco Javier Iglesias-Rodríguez. Application of an svm-based regression model to the air quality study at local scale in the avilés urban area (spain). *Mathematical and Computer Modelling*, 54(5-6):1453–1466, 2011.

[173] Antonio Bernardo Sánchez, Celestino Ordóñez, Fernando Sánchez Lasheras, Francisco Javier de Cos Juez, and Javier Roca-Pardiñas. Forecasting so2 pollution incidents by means of elman artificial neural networks and arima models. In *Abstract and applied analysis*, volume 2013, page 238259. Wiley Online Library, 2013.

[174] ML Sanchez, D Pascual, C Ramos, and I Perez. Forecasting particulate pollutant concentrations in a city from meteorological variables and regional weather patterns. *Atmospheric Environment. Part A. General Topics*, 24(6):1509–1519, 1990.

[175] Youngguk Seo and Seong-Min Kim. Estimation of greenhouse gas emissions from road traffic: A case study in korea. *Renewable and Sustainable Energy Reviews*, 28:777–787, 2013.

[176] Seyedeh Reyhaneh Shams, Ali Jahani, Saba Kalantary, Mazaher Moeinaddini, and Nematollah Khorasani. Artificial intelligence accuracy assessment in no2 concentration forecasting of metropolises air. *Scientific Reports*, 11(1):1805, 2021.

[177] Seyedeh Reyhaneh Shams, Ali Jahani, Saba Kalantary, Mazaher Moeinaddini, and Nematollah Khorasani. The evaluation on artificial neural networks (ann) and multiple linear regressions (mlr) models for predicting so2 concentration. *Urban Climate*, 37:100837, 2021.

[178] Samuel S Shapiro, Martin B Wilk, and Hwei J Chen. A comparative study of various tests for normality. *Journal of the American statistical association*, 63(324):1343–1372, 1968.

[179] Samuel Sanford Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.

[180] Ekta Sharma, Ravinesh C Deo, Ramendra Prasad, Alfio V Parisi, and Nawin Raj. Deep air quality forecasts: suspended particulate matter modeling with convolutional neural and long short-term memory networks. *Ieee Access*, 8:209503–209516, 2020.

[181] Moolchand Sharma, Samyak Jain, Sidhant Mittal, and Tariq Hussain Sheikh. Forecasting and prediction of air pollutants concentrates using machine learning techniques: the case of india. In *IOP conference series: Materials science and engineering*, volume 1022, page 012123. IOP Publishing, 2021.

[182] Justin Shen, Davesh Valagolam, and Serena McCalla. Prophet forecasting model: A machine learning approach to predict the concentration of air pollutants (pm2. 5, pm10, o3, no2, so2, co) in seoul, south korea. *PeerJ*, 8:e9961, 2020.

[183] Pierre Sicard, Paola Crippa, Alessandra De Marco, Stefano Castruccio, Paolo Giani, Juan Cuesta, Elena Paoletti, Zhaozhong Feng, and Alessandro Anav. High spatial resolution wrf-chem model over asia: Physics and chemistry evaluation. *Atmospheric Environment*, 244:118004, 2021.

[184] Zbigniew Skrobacki, Lidia Ewa Dabek, and Piotr Lagowski. Modelling and forecasting of so2 concentration in atmospheric air–a case study of the city of krakow. In *IOP conference series: materials science and engineering*, volume 471, page 102057. IOP Publishing, 2019.

[185] Sang Hyun Sohn, Sea Cheon Oh, and Yeong-Koo Yeo. Prediction of air pollutants by using an artificial neural network. *Korean Journal of Chemical Engineering*, 16:382–387, 1999.

[186] Chao Song and Xiaoshuang Fu. Research on different weight combination in air quality forecasting models. *Journal of Cleaner Production*, 261:121169, 2020.

[187] Izabela Sówka, Anna Chlebowska-Styś, Łukasz Pachurka, Wioletta Rogula-Kozłowska, and Barbara Mathews. Analysis of particulate matter concentration variability and origin in selected urban areas in poland. *Sustainability*, 11(20):5735, 2019.

[188] Shankar Subramaniam, Naveenkumar Raju, Abbas Ganesan, Nithyaprakash Rajavel, Maheswari Chenniappan, Chander Prakash, Alokesh Pramanik, Animesh Kumar Basak, and Saurav Dixit. Artificial intelligence technologies for forecasting air pollution and human health: A narrative review. *Sustainability*, 14(16), 2022.

[189] Shan Suthaharan. Deep learning models. In *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*, pages 289–307. Springer, 2016.

[190] Prabhat K Swamee and Aditya Tyagi. Formation of an air pollution index. *Journal of the Air & Waste Management Association*, 49(1):88–91, 1999.

[191] Kimberly Tasha Jiayi Tang, Changqing Lin, Zhe Wang, Sik Wing Pang, Tze-Wai Wong, Ignatius Tak Sun Yu, Wallace Wai Yip Fung, Md Shakhaoat Hossain, and Alexis K.H. Lau. Update of air quality health index (aqhi) and harmonization of health protection and climate mitigation. *Atmospheric Environment*, 326:120473, 2024.

[192] Fernando Terroso-Saenz, Juan Morales-García, and Andres Muñoz. Nationwide air pollution forecasting with heterogeneous graph neural networks. *ACM Trans. Intell. Syst. Technol.*, 15(1), January 2024.

[193] Min Ye Thu, Wunna Htun, Yan Lin Aung, Pyone Ei Ei Shwe, and Nay Min Tun. Smart air quality monitoring system with lorawan. In *2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS)*, pages 10–15. IEEE, 2018.

[194] Yi-Ting Tsai, Yu-Ren Zeng, and Yue-Shan Chang. Air pollution forecasting using rnn with lstm. In *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, pages 1074–1079. IEEE, 2018.

[195] Richard E Turner. An introduction to transformers. *arXiv preprint arXiv:2304.10557*, 2023.

[196] Mihaela T Udristioiu, Youness EL Mghouchi, and Hasan Yildizhan. Prediction, modelling, and forecasting of pm and aqi using hybrid machine learning. *Journal of Cleaner Production*, 421:138496, 2023.

[197] A Usha Ruby, J George Chellin Chandran, Prasannavenkatesan Theerthagiri, Renuka Patil, BN Chaithanya, and TJ Swasthika Jain. Forecasting pm2. 5 concentration using gradient-boosted regression tree with cnn learning model. *Optical Memory and Neural Networks*, 33(1):86–96, 2024.

[198] Sebastien Pérez Vasseur and José L Aznarte. Comparing quantile regression methods for probabilistic forecasting of no2 pollution levels. *Scientific Reports*, 11(1):11592, 2021.

[199] A Vlachogianni, P Kassomenos, Ari Karppinen, S Karakitsios, and Jaakko Kukkonen. Evaluation of a multiple regression model for the forecasting of the concentrations of nox and pm10 in athens and helsinki. *Science of the total environment*, 409(8):1559–1571, 2011.

[200] Bin Wang, Deqing Yi, Lei Jin, Zhiwen Li, Jufen Liu, Yali Zhang, Xinghua Qiu, Wenxin Liu, Shu Tao, and Aiguo Ren. Organochlorine pesticide levels in maternal serum and risk of neural tube defects in offspring in shanxi province, china: A case–control study. *Science of The Total Environment*, 490:1037–1043, 2014.

[201] Gang Wang, Chunlai Yu, Kai Guo, Haisong Guo, and Yibo Wang. Research of low-cost air quality monitoring models with different machine learning algorithms. *Atmospheric Measurement Techniques*, 17(1):181–196, 2024.

[202] Jianzhou Wang, Tong Niu, and Rui Wang. Research and application of an air quality early warning system based on a modified least squares support vector machine and a cloud model. *International journal of environmental research and public health*, 14(3):249, 2017.

[203] Jingyang Wang, Xiaolei Li, Lukai Jin, Jiazheng Li, Qiuhong Sun, and Haiyao Wang. An air quality index prediction model based on cnn-ilstm. *Scientific Reports*, 12(1):8373, 2022.

[204] Xiaohu Wang, Suo Zhang, Yi Chen, Longying He, Yongmei Ren, Zhen Zhang, Juan Li, and Shiqing Zhang. Air quality forecasting using a spatiotemporal hybrid deep learning model based on vmd–gat–bilstm. *Scientific Reports*, 14(1):17841, 2024.

[205] Qingsong Wen, Tian Zhou, Chaoli Zhang, Weiqi Chen, Ziqing Ma, Junchi Yan, and Liang Sun. Transformers in time series: A survey. *arXiv preprint arXiv:2202.07125*, 2022.

[206] Will Williams, Sam Ringer, Tom Ash, David MacLeod, Jamie Dougherty, and John Hughes. Hierarchical quantized autoencoders. *Advances in Neural Information Processing Systems*, 33:4524–4535, 2020.

[207] Cort J Willmott and Kenji Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82, 2005.

[208] Kinga Wiśniewska, Anita Urszula Lewandowska, and Marta Staniszewska. Air quality at two stations (gdynia and rumia) located in the region of gulf of gdansk during periods of intensive smog in poland. *Air Quality, Atmosphere & Health*, 12(7):879–890, 2019.

[209] Wan-Sik Won, Rosy Oh, Woojoo Lee, Sungkwan Ku, Pei-Chen Su, and Yong-Jin Yoon. Hygroscopic properties of particulate matter and effects of their interactions with weather on visibility. *Scientific reports*, 11(1):16401, 2021.

[210] Jin-Zhun Wu, Dan-Dan Ge, Lin-Fu Zhou, Ling-Yun Hou, Ying Zhou, and Qi-Yuan Li. Effects of particulate matter on allergic respiratory diseases. *Chronic Diseases and Translational Medicine*, 4(2):95–102, 2018. Special Issue: Air Pollution and Chronic Respiratory Diseases.

[211] Barry J Wythoff. Backpropagation neural networks: a tutorial. *Chemometrics and Intelligent Laboratory Systems*, 18(2):115–155, 1993.

[212] Jiahui Xu, Ling Chen, Mingqi Lv, Chaoqun Zhan, Sanjian Chen, and Jian Chang. Highair: A hierarchical graph neural network-based air quality forecasting method. *arXiv preprint arXiv:2101.04264*, 2021.

[213] Rui Xu, Dawen Yao, Yuzhuang Pian, Ruhui Cao, Yixin Fu, Xinru Yang, Ting Gan, and Yonghong Liu. Integrating mobile and fixed monitoring data for high-resolution pm2. 5 mapping using machine learning. *arXiv preprint arXiv:2503.12367*, 2025.

[214] Kalpit Yadav, Vipul Arora, Mohit Kumar, Sachchida Nand Tripathi, Vidyanand Motiram Motghare, and Karansingh A Rajput. Few-shot calibration of low-cost air pollution (pm _{2.5}) sensors using meta learning. *IEEE Sensors Letters*, 6(5):1–4, 2022.

[215] Rui Yan, Jiaqiang Liao, Jie Yang, Wei Sun, Mingyue Nong, and Feipeng Li. Multi-hour and multi-site air quality index forecasting in beijing using cnn, lstm, cnn-lstm, and spatiotemporal clustering. *Expert Systems with Applications*, 169:114513, 2021.

[216] Hufang Yang, Zhijie Zhu, Chen Li, and Ranran Li. A novel combined forecasting system for air pollutants concentration based on fuzzy theory and optimization of aggregation weight. *Applied Soft Computing*, 87:105972, 2020.

[217] Chunyong Yin, Sun Zhang, Jin Wang, and Neal N Xiong. Anomaly detection based on convolutional recurrent autoencoder for iot time series. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(1):112–122, 2020.

[218] Aynur Yonar and Harun Yonar. Modeling air pollution by integrating anfis and metaheuristic algorithms. *Modeling Earth Systems and Environment*, 9(2):1621–1631, 2023.

[219] Abolfazl Younesi, Mohsen Ansari, Mohammadamin Fazli, Alireza Ejlali, Muhammad Shafique, and Jörg Henkel. A comprehensive survey of convolutions in deep learning: Applications, challenges, and future trends. *IEEE Access*, 12:41180–41218, 2024.

[220] Chengqing Yu, Fei Wang, Yilun Wang, Zezhi Shao, Tao Sun, Di Yao, and Yongjun Xu. Mgsfformer: A multi-granularity spatiotemporal fusion transformer for air quality prediction. *Information Fusion*, 113:102607, 2025.

[221] Ruiyun Yu, Yu Yang, Leyou Yang, Guangjie Han, and Oguti Ann Move. Raq–a random forest approach for predicting air quality in urban sensing systems. *Sensors*, 16(1):86, 2016.

[222] Jaeseok Yun and Jiyoung Woo. Iot-enabled particulate matter monitoring and forecasting method based on cluster analysis. *IEEE Internet of Things Journal*, 8(9):7380–7393, 2020.

[223] Adrian Zafra-Pérez, Jonathan Medina-García, C. Boente, Juan Antonio Gómez-Galán, Ana Maria Sánchez de la Campa, and Jesus D. de la Rosa. Designing a low-cost wireless sensor network for particulate matter monitoring: Implementation, calibration, and field-test. *Atmospheric Pollution Research*, 15(9):102208, 2024.

[224] Mateusz Zareba, Szymon Cogiel, and Tomasz Danek. Spatio-temporal pm2. 5 forecasting using machine learning and low-cost sensors: An urban perspective. *Engineering Proceedings*, 101(1):6, 2025.

[225] Masoomeh Zeinalnezhad, Abdoulmohammad Gholamzadeh Chofreh, Feybi Ariani Goni, and Jiří Jaromír Klemeš. Air pollution prediction using semi-experimental regression model and adaptive neuro-fuzzy inference system. *Journal of Cleaner Production*, 261:121218, 2020.

[226] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 11121–11128, 2023.

[227] Lanyi Zhang, Jane Lin, Rongzu Qiu, Xisheng Hu, Huihui Zhang, Qingyao Chen, Huamei Tan, Danting Lin, and Jiankai Wang. Trend analysis and forecast of pm2. 5 in fuzhou, china using the arima model. *Ecological indicators*, 95:702–710, 2018.

[228] Dixian Zhu, Changjie Cai, Tianbao Yang, and Xun Zhou. A machine learning approach for air quality prediction: Model regularization and optimization. *Big data and cognitive computing*, 2(1):5, 2018.

[229] Suling Zhu, Xuanlin Qiu, Yanru Yin, Min Fang, Xingrong Liu, Xuejing Zhao, and Yanjun Shi. Two-step-hybrid model based on data preprocessing and intelligent optimization algorithms (cs and gwo) for no2 and so2 forecasting. *Atmospheric Pollution Research*, 10(4):1326–1335, 2019.

[230] Naomi Zimmerman, Albert A Presto, Sriniwasa PN Kumar, Jason Gu, Aliaksei Hauryliuk, Ellis S Robinson, Allen L Robinson, and Ramachandran Subramanian. A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmospheric Measurement Techniques*, 11(1):291–313, 2018.