Politechnika Wrocławska

FIELD OF SCIENCE: FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
DISCIPLINE OF SCIENCE: INFORMATION AND COMMUNICATION TECHNOLOGY

DOCTORAL DISSERTATION

Faults diagnosis in cellular radio networks and their coverage issues detection with advanced analysis of mobile networks' metrics

Malgorzata Tomala

Supervisor: prof. dr hab. inż. Kamil Staniec

Keywords: ARTIFICIAL INTELLIGENCE, MACHINE LEARNING, 5G, RADIO ACCESS NETWORK, RADIO INTERFACE, DRIVE TESTS, SELF OPTIMIZING NETWORKS, RADIO RESOURCE CONTROL

WROCŁAW 2024

Streszczenie

Piąta generacja (5G) sieci komórkowych jest rozpatrywana jako ewolucja społeczno-techniczna poprzez umożliwienie ścisłej ingerencji i integracji komunikacji mobilnej z codziennym życiem całego społeczeństwa. Wynika to z przewidywanych możliwości sieci 5G, takich jak bezprecedensowe poziomy przepustowości danych, ultra niezawodna komunikacja o niskich opóźnieniach oraz masowa łączność urządzeń. Wsparcie idących za tym ulepszeń komunikacyjnych, które wcześniej były niewyobrażalne, nakłada również nowe wymagania projektowe na systemy 5G. Sprostanie tak szerokiej gamie wymagań, które często stoją w opozycji do siebie, nie jest łatwe. Stworzenie efektywnego ekosystemu 5G wymaga starannych analiz i innowacyjnych rozwiązań, aby umożliwić zharmonizowanie zróżnicowanych potrzeb operacyjne w jednej sieci. Ze względu dalekosiężne obietnice, sztuczna inteligencja (ang. Artificial Intelligence, AI) jest postrzegana jako mająca ogromny potencjał, który będzie można wykorzystać do rozwiązywania tychże wyzwań stojących przed projektantami sieci telekomunikacyjnych. W rzeczywistości rozwój badawczy w dziedzinach AI i 5G przenika się wzajemnie i pojawienie się tych trendów jednocześnie jest kluczową inspiracją tej rozprawy. Przeprowadzona w ramach niniejszej pracy geneza istniejących wyzwań w procesach automatyzacji i optymalizacji sieci 5G, z którymi trzeba się zmierzyć, aby zapewnić wysoką wydajność sieci telekomunikacyjnych, skupiła pytanie badawcze na tym: w jaki sposób zastosować sztuczną inteligencję w sieci dostępu radiowego 5G (ang. Radio Access Network RAN) oraz jak zrealizować dostarczanie parametrów wejściowych algorytmu uczenia maszynowego (ML), automatycznym gromadzeniem kluczowych danych pochodzących z interfejsu radiowego, przy jednoczesnym skutecznym zmniejszeniu ilości przesyłu danych. W tym celu, pierwszą czynnością w ramach badania, było przeprowadzenie inspekcji podstawowego typu ruchu 5G (a konkretnie ulepszonej mobilnej łączności szerokopasmowej, ang. Enhanced Mobile Broadband eMBB) w rzeczywiście wdrożonej sieci 5G typu NSA (ang. Non-Stand Alone), poprzez manualne zbiory danych z komercyjnych telefonów 5G. Serie danych, będące krytycznymi wskaźnikami wydajności radiowej, zostały poddane statystycznym analizom, które ujawniły konkretne trendy i relacje między zebranymi parametrami, oraz nakreśliły ich znaczenie w sieci RAN. Wyniki te pomogły następnie w doborze algorytmów uczenia maszynowego oraz w identyfikacji parametrów niezbędnych do monitorowania danych a także wskaźników, które powinny stanowić istotny cel predykcji podczas badania interfejsu radiowego. Podczas kolejnego kroku badawczego: trenowania i testowania trzech obiecujących algorytmów nadzorowanego uczenia maszynowego (ang. Supervised Machine Learning), tylko algorytm 'Decision Tree' z nieliniową i zaawansowaną zdolnością regresji został zweryfikowany z sukcesem, poprzez wykazanie najwyższej wydajności. W celu dopełnienia metody badawczej w rozprawie zaproponowano schemat monitorowania urządzeń komercyjnych w sieci 5G, poprzez konfigurowanie ich procedurą automatycznego zbierania danych, inicjowaną ze stacji bazowej (gNB) oraz opartą o ewaluację działania adekwatnego algorytmu ML. Procedura, umożliwia dostosowywanie interwałów raportowania metryk, w oparciu o odpowiedni wynik algorytmu, co zapewnia lepszą skuteczność w porównaniu do istniejących metod, poprzez ograniczenie masowego gromadzenia danych w sieci dostępowej RAN. Dodatkowo, przedstawiona technika prezentuje, jak zestaw wskaźników wejściowych może zostać zredukowany i ograniczony tylko do jednego najbardziej krytycznego parametru (tj. mocy odbieranego sygnału), co umożliwia dalszą redukcję i uzyskanie wyższej wydajności w sposobie gromadzenia danych.

Podsumowując, praca badawcza potwierdza, że automatyczny proces monitorowania wybranych charakterystyk radiowych, ulepszony o właściwości uczenia się i przewidywania opartego na sztucznej inteligencji, jest realną alternatywą względem manualnych metod gromadzenia danych (np. poprzez tradycyjne „drive-testy") oraz przynosi wymierne korzyści w postaci kontroli i/lub redukcji ilości monitorowanych danych. Proponowana metoda wykazuje łatwość adaptacji, ponieważ wykorzystuje w dużej mierze istniejące elementy i procedury sieci 5G, wprowadzając algorytm ML, jako ulepszenie funkcjonujących metod monitorowania. Zaproponowaną metodę można następnie skalować, w celu poszerzenia zakresu użyteczności, poprzez poszerzenie zbiorów danych o inny

rodzaj informacji. Ponadto, badania podkreślają dużą elastyczność opracowanej techniki uczenia maszynowego poprzez przedstawienie możliwości adaptacji w dwóch elementach sieci radiowej: jako modelu uczenia maszynowego po stronie stacji bazowej (gNB) lub modelu uczenia maszynowego po stronie terminala (UE). Tym samym, dowiedziono praktycznej wykonalności integracji sztucznej inteligencji z siecią dostępową 5G (RAN), oraz zademonstrowano strategie, które można zastosować w celu skutecznego wdrożenia sztucznej inteligencji w sieci 5G.

Abstract

Fifth Generation (5G) of mobile networks has been envisioned to initiate socio-technical evolution through mobile communications becoming closely integrated in the daily life of the whole society. This results from enabling unprecedented levels of data throughput, ultra-reliable low-latency communication, and massive device connectivity. Fostering previously unimaginable advancements imposes also requirements to 5G system design. To meet a wide variety of demands, often contradicting, is a challenge. Creating an effective 5G ecosystem demands careful consideration and innovative solutions to harmonize the diverse operational needs. Due to strong-fitting capabilities, Artificial Intelligence (AI) is seen as a great potential, that can be employed to solve challenges faced in telecommunication networks. In fact, AI and 5G are permeating incrementally and the emergence of both is the key inspiration for this dissertation. Within the scope of this work, challenges in 5G network optimization processes are elaborated, addressing various complex and interrelated problems that are faced to ensure high-performing telecommunication networks. With a focus to provide reliability in the network operations originating from radio access, the work addresses the question: how to apply Artificial Intelligence in 5G Radio Access Network (RAN) and feed Machine Learning (ML) techniques with the radio characteristic-based automatic data collection. For this purpose, the research conducts the streamlined 5G use case traffic (namely enhanced Mobile Broadband(eMBB)) inspection in real 5G Non-Stand Alone deployment. To grasp practical effects of the 5G network on end user performance, in various deployment scenarios, the data has been collected from commercial 5G smartphones The analyses of data series uncover trends and relationships among radio performance indicators and recognize their varying importance in RAN. The findings further guide ML algorithm selection, including identification of essential parameters for data monitoring, that prove to be adequate target for predictions, when examining radio interface. Throughout training and testing three candidate Supervised Machine Learning algorithms with the collected datasets, Decision Tree algorithm with non-linear and advanced regression capability is validated, by showing most efficient performance. To address the research question on how the ML algorithm can be integrated into RAN operation, the dissertation further proposes an effective monitoring scheme to involve commercial devices and configure them with metrics scheduled by ML-empowered automated data collection procedure by gNB. The procedure enables metrics' reporting intervals adjustments, based on the suitable algorithm outcome. This ensures effectiveness by limiting massive data collection in RAN. Additionally, the set of input metrics can be reduced and limited to just one critical parameter (i.e. received signal strength), enabling intelligent optimization of the data collection and better performance, once compared to traditional methods on data collection. In conclusion, the research validates, that replacing drive tests, with automated data collection of a few selected radio characteristics, with operational integration of learning and prediction capabilities, becomes a viable ML-enabled monitoring framework. The introduced method brings measurable benefits in the form of control and/or reduction of the amount of monitored data. Furthermore, it can be easily adopted, as integrates existing measurement collection tools with ML algorithm tailored to advance the functioning methods in place. The method can scale to address a wider range of use cases, by extension of the datasets with other type of information. Moreover, the research underscores the developed ML technique great flexibility by marking adoption capabilities to two possible approaches: the gNB-sided ML model and the UE-sided ML model, with different operational and implementational impacts. Consequently, it proves practical viability of AI integration into 5G RAN, showcases feasibility and demonstrates advantageous strategies that can be aimed for successful uptake of AI in 5G RAN.

Table of contents

1 Introduction ... 7
1.1 5G Telecommunication Networks ... 10
1.1.1 System Architecture ... 10
1.1.2 5G Air Interface ... 12
1.1.3 Key 5G Traffic Use Cases ... 17
1.1.4 5G Network Performance Optimization Methods ... 19
1.2 AI/ML Philosophy and Development ... 23
1.2.1 Power of AI ... 23
1.2.2 AI Categories ... 24
1.2.3 Machine Learning ... 25
1.2.4 Machine Learning Categories ... 26
1.2.5 Artificial Neural Networks and Deep Learning ... 27
1.2.6 Machine Learning Core Algorithms Overview ... 28
2 Research goal ... 30
2.1 Research problem ... 30
2.2 Research questions ... 31
2.3 Objective and strategy ... 31
2.4 Theses ... 32
3 Literature overview ... 33
3.1 Network Performance Monitoring ... 33
3.1.1 3GPP standard developments ... 33
3.1.2 Research developments ... 38
3.2 ML Applicability to Mobile Networks ... 41
3.2.1 3GPP standard developments ... 41
3.2.2 Research developments ... 44
3.3 Literature overview conclusion ... 49
3.4 Identified cognitive aps in the literature ... 49
4 Methodology ... 50
4.1 Methodology Design ... 50
4.2 Experimental Data ... 51
4.2.1 Data Collection ... 51
4.2.2 Drive Test Routes ... 53

4.2.3 Data Samples ... 56
4.2.4 Data Limitations ... 58
4.3 Data Analysis ... 58
4.3.1 Trends in parameters ... 58
4.3.2 Results analysis ... 60
4.3.2.1 Results for Urban area ... 60
4.3.2.2 Results for Sub-urban and Rural area ... 64
4.3.2.3 Results for Highway area ... 68
4.3.2.4 Drive Test results summary ... 70
4.3.3 Trends in correlation ... 71
4.3.4 Correlation analysis ... 71
4.3.4.1 Results for Urban area ... 72
4.3.4.2 Results for Sub-urban area ... 74
4.3.4.1 Results for Urban area: additional parameters ... 75
4.3.4.2 Correlation analysis summary ... 76
4.3.5 Data analysis summary ... 76
4.4 ML application ... 77
4.4.1 ML algorithm selection ... 78
4.4.1.1 Linear regression algorithm ... 78
4.4.1.2 Decision tree algorithm ... 79
4.4.1.3 Support Vector Machine algorithm ... 79
4.4.2 ML model architecture ... 80
4.4.3 ML model performance ... 81
4.4.3.1 Results for Urban area ... 82
4.4.3.2 Results for Sub-urban area ... 83
4.4.4 Findings on ML application ... 86
5 Strategy to implement ... 88
5.1 Foundations ... 88
5.2 RAN-sided ML model ... 91
5.3 UE-sided ML model ... 92
5.4 Benefits ... 93
6 Conclusion ... 97
List of tables ... 102
List of figures ... 102
References ... 104

1 Introduction

With the rapid advancement of wireless networks, the ways of communication have undergone a significant transformation and include several technological breakthroughs. The first generation of mobile telecommunication (1G) was based on analog transmission for voice, the second generation of mobile telecommunication (2G) networks introduced digital communications, with data and short messaging services for mobile devices. The third generation of mobile telephony (3G) evolved for faster data transfers and improved voice quality, driving mass adoption of mobile broadband. The decade from 2010 to 2020, dedicated to fourth generation of mobile telecommunication (4G) technology, has enabled further increase of the speed and capacity of cellular networks. It initiated the rise to an entirely new generation of applications, which fundamentally changed how mobile networks connect people. With 4G it became easier than ever to share and access information digitally, which accelerated the telecommunications industry rise in popularity of mobile phones and high-speed internet access. These days, in the advent of 5G telecommunication networks, a significant leap from the previous generations is marked in terms of speed, connectivity, and functionality. The fifth generation of mobile telecommunication (5G) technology vision brings with improvements in many ways — up to 10 times faster speed than 4G, up to 10 times lower latency than 4G, up to 100 times as many devices as 4G, and up to 90 percent lower energy consumption compared to 4G [1]. By enabling unprecedented levels of data throughput, ultra-reliable low-latency communication, and massive device connectivity, 5G has been envisioned to revolutionize various industries. From advanced, immersive real-time video communication, smart homes and environments to autonomous vehicles and remote healthcare, the usage requirements have been set to enhance the quality of life by providing seamless connectivity and fostering advancements that were previously unimaginable. The foreseen transformation has been not just about higher speeds but also socio-technical evolution through mobile communications becoming closely integrated in the daily life of the whole society. The earliest technological solutions in mobile networks aimed at enabling wireless connectivity transfer for cellular data voice mainly. The 5G telecommunication network is hailed as digitalization and connection of everything and everyone with the goal of automating much of life [4], [5].

Today, 5G deployments are in a full swing [2],[3]. Blind spots remain, but 95 per cent of the world's population is now living within range of a mobile broadband network and more than 90 per cent of the population own a mobile phone, which is considered as the most common gateway to the mobile broadband network access. In result, the data volumes in mobile networks have started greatly exceeding the previous generations' levels. According to the official figures, global mobile-broadband traffic rates are estimated to have already reached nearly 1k exabytes[1] yearly per landmass [4].

Notably, the connected society and entities, that are bound to have varying levels of critical requirements, and performance indices need to be accommodated in a collaborative system operation. The supporting mobile network systems need to handle the boosted data rates and diverse usage scenarios, simultaneously. International Telecommunication Union (ITU), working towards providing successful telecommunication network deployments at the regional and international levels, established high level framework and overall objectives for 5G developments in [6],[7]. The approaches for viable systems improve spectral efficiency by allocating more spectrum resources and exploiting larger bandwidths for the next generation of wireless communication networks [7].

Besides the undergoing data traffic increase and changes in advanced deployments techniques, end user expectations remain constant: network subscribers expect a satisfactory end-user experience. To offer the best experience to any 5G terminal, robust and reliable connectivity solutions are needed as well as the ability to efficiently maintain service quality. Mobile operators and network vendors enrolling 5G, face far more demanding challenges to maintain the service at a satisfactory level.

---

[1] 1 Exabyte = 10^{18} bytes. Global mobile-broadband traffic rates were estimated to have reached 913 exabytes (EB) in 2022

There are two relevant frameworks facilitating network service management and optimization. Self-Organizing Network (SON) technology is one solution that allows operators to improve and enhance network efficiency. This technology automates network operations through self-commissioning (RAN node plug and play), self-healing, and self-configuration. These functions enable autonomous configuration and monitoring, network parameters tuning and appropriate steering of the network traffic and load, resulting in improved overall network performance [8]. Another example is Minimization of Drive Tests (MDT), which is a method that utilize users' equipment (UE) to collect radio measurements and associated location information. The collected data can be used to evaluate network performance while reducing the Operational Expenditure (OPEX) associated with traditional manual drive tests. MDT allows operators to collect feedback from regular users' devices in the field (including indoor users) without a need to conduct trial data collection sessions the (before actual network rollout) to observe network performance in a specific area [9]. SON focuses on the operational efficiency of the network. MDT is designed to collect actual data in a real time operation and improving end-user experience based on offline data post-processing and assessment. These two features are complementary, addressing major network optimization use cases. When applied together, they synergize to achieve better network performance and satisfied subscribers.

These methods at the very best may only serve the baseline to acquire necessary feedback from the users and apply certain optimization strategy or corrective action per designed optimization use case after certain observability period. However, they are not provisioned with a real time intelligent forecasting and decision-making strategies. For that reason, rapidly developing Artificial Intelligence (AI) becomes attractive strength. In fact, the intelligence is a cornerstone of AI domain, which has become of particular interest in telecom sector. The telecommunication network's traffic, in context of transferred data or management and processing related information, generates vast amounts of data daily. By design and its purpose, AI can process and analyze big data to derive valuable insights, improve decision-making, and optimize data collection as well as its processing. AI has stunned research and technology industry due to its potential to revolutionize the sector, as it is seen as an innovative way to embrace automated solutions for network performance optimization and thus promote a sustainable business [10].

Automated approaches exist, however 5G era advancements causing a transforming paradigm shift in the cellular communications industry appears much more complex due to the 5G traffic variation characteristic of different users, flexible network deployments, as well as varying implementations and capabilities. Overall, coexistence of flexible deployments and connections options and the need to process very high volumes of data become a bottleneck for the effectiveness and efficiency of the known autonomous solutions [11]. Therefore, AI-based investigation in telecom refers to the heightened excitement and anticipation surrounding the potential of AI to comprehend and support the undergoing changes in telecommunications industry.

The intensified research and growing importance of AI applicability to mobile networks is driven by several emerging factors, including technological advancements, market demands, and the promises of AI to solve long-standing challenges in telecom. The key drivers result from strong fitting AI capabilities to 5G technological evolution: Machine Learning (ML) and Deep Learning (DL) as AI subfields have shown impressive results in data analysis, pattern recognition, and predictive abilities. These methods are also highly applicable to telecom. Besides the fundamental need to treat series of big data, the deployment of 5G networks and edge computing provides the necessary infrastructure to support AI applications, enabling real-time data processing and low-latency responses. In addition, there is a need for more efficient and reliable network management due to market demands evolving towards increased connectivity, through the rapid uptake of smartphones, massive proliferation of devices, including Internet of Things (IoT) and high-bandwidth applications. At the same time, there is a need to satisfy enhanced customer experience, for which AI promises to provide supercharged opportunities, improved abilities on data analysis and thus predict or even resolve issues before they affect users. Yet, AI can optimize network resources, manage traffic, and reduce latency, improving overall network efficiency and user experience. By analyzing data from different network components, AI can predict failures and enable maintenance in proactive manner, reducing downtime and operational costs, ensuring predictive maintenance. Furthermore, AI can automate routine tasks such as network configuration, monitoring, and fault management, reducing human involvement and operational costs [12].

Numerous organizations within the telecom sector acknowledge that the potential benefits presented by AI are matched, if not surpassed [13]. The magnitude of this trend is also reflected by standard development organizations doings. 3GPP (3rd Group Partnership Project) standard, which gathers major players and ensures the industry alignment for mobile networks, has undertaken the mission to integrate AI solutions into 5G Advanced technology. The standardization consortium has set the official roadmap for 5G system architecture to be explicitly associated with AI [14]. The work is in progress, but the investigations have clear vision: to define key building blocks for AI-enabled mobile networks. The standard's specifications support is aimed to be ranging from identifying generic requirements that AI-based methods should comply with, once applied in wireless networks, through principles of Machine Learning (ML) use depending on different scenarios to prominent management operations. Specifically, it targets definition of a functional framework that will enable AI utilization in 5G system and identification of the architecture i.e.: network entities and protocols involved in data acquisition and exposure. This includes the segment of Radio Access Network (RAN), where the AI applicability depends on 5G use cases severity.

The standard-driven exploration is a milestone, proving AI and 5G are permeating incrementally with unavoidable societal impact. Indeed, the trends within the telecommunication standards show that there is a growing reliance on AI to optimize services, with the potential to further increase profitability and operational effectiveness. Nevertheless, as the ongoing efforts do not necessitate proving anticipated ML algorithms gains [15], the choice for actual ML algorithm choice is left to communication services providers. Given diverse characteristics of well exploited 5G traffic types [7], the obtained results determined by different ML algorithms use, can potentially diversify outcomes. Even by following the guided framework, the introduction of AI-driven techniques can take place in a variety of ways. The approaches can range from more uniform to truly and explicitly associated with a specific use case. Without exercising the AI utilization with specific ML algorithms [15], useful development of technical solutions and practical implementations remains uncertain. Such indeterminacy can be seen as another-biggest challenge (next to complexity, investment costs or difficulty of integrating with existing tools/infrastructure), preventing organizations from investing in tools for AI operation.

This work explores how to apply AI strategy guided by 3GPP standard development organization and execute the principles in 5G RAN network deployments. It examinates how to initiate and facilitate AI algorithm-based optimization based on essential radio measurements that are anyway made available in the network as provided by the commercial devices for regular radio link monitoring purposes. The work validates benefits of the use of AI in 5G mobile networks, by qualifying of adequate ML algorithms applicability through real 5G traffic observability. The main contribution is to intelligently analyze the data with strong generalization abilities, within standardized 5G architecture and framework. Also, thanks to strong generalization abilities regardless of demanding and diversified 5G traffic types, it provides a better understanding of the potential gains and associated complexity avoidance.

This effort prioritizes augmenting the air-interface monitoring with support of AI algorithms for overall enhanced performance and reduced overhead, resulting in a promising forecast for future air-interface use cases leveraging AI techniques, with various levels of collaboration between the gNB and UE.

To focus the study through the ML algorithms integration perspective, it is organized in the following manner:

- Chapter 1 presents background to the two subject matter domains: 5G networks and AI philosophy, including: introduction to 5G system foundations and architecture, basics of existing network optimization automated tools, all based on developments by 3GPP standards; a generic description of AI tools, ML approaches, and ML algorithms classification.
- Chapter 2 identifies research questions (Chapter 2.2).and states theses guiding this work (Chapter 2.4).
- Chapter 3 presents current state of the art, through research articles and standardization sources, insight into academic research on explorations AI applicability in various layers and areas of mobile

network, seeking new insights and understandings on ML usefulness in complex systems as wireless telecommunication.

- Chapter 4 introduces methodology undertaken to prove the theses, presents experimental data collection (real measurements recorded by the end users in 5G network), provides detailed analysis on the key radio characteristics in Chapter 4.3, differentiating the metrics recorded by the end users in varying scenarios (from rural to urban scenarios), and in Chapter 4.4 provides experimental analysis on candidate ML algorithms applied on the data, validating good performance of the most suitable one.
- Chapter 5 presents execution strategy of implementing the experiment outcome in 5G network, with two approaches for ML algorithm adoption (either at the base station or user device), and proves operational benefits.
- Chapter 6 concludes the work and guides further development of technical solutions and practical implementations, listing prospective challenges of shaping the future of AI integration in telecom.

As a whole, this work explores the application of Artificial Intelligence in 5G Radio Access Network. It provides in-depth guide on how to utilize Machine Learning techniques by leveraging radio characteristic-based automatic data collection to evaluate 5G performance. The examination endorses ML tools for the 5G portfolio scenarios, validates their usefulness and gains in accessing the requirements of various 5G portfolio scenarios. The proposed methodology builds on existing network infrastructure and thus facilitate the ML-driven data monitoring into established network performance optimization methods, including Self-Organizing Networks (SON) and Minimization of Drive Tests (MDT). This methodology provides guidance for future network deployments and implementations adopted on 3GPP standard basis.

1.1 5G Telecommunication Networks

1.1.1 System Architecture

The fifth-generation (5G) of mobile communications system has been defined by 3GPP standards development organization since its release 15 versions of the technical specifications (TS) [16],[17][18],[19],[20]. 5G has been designed to meet the performance requirements set by the International Telecommunication Union (ITU) for International Mobile Telecommunications for 2020 (IMT-2020) and beyond [21],[22]. Recommendations for the systems were assuming harmonized use of the existing IMT components. Therefore, an initial development strategy for 5G was to utilize Long Term Evolution (LTE) base stations for initial access and mobility handling, to enable reuse of the existing network infrastructure. This enabled to account for non stand-alone (NSA) architecture options such as connecting 5G NodeB (gNB) to the Evolved Packet Core (EPC) and operating NR and LTE in multi-connectivity mode with NR as the master node and LTE as the secondary node.

While the preceding technological solutions in mobile networks (including 4G) primarily aimed at enabling wireless connectivity for cellular voice data, the 5G network has been seen as a revolutionary milestone for its potential to digitalize and connect everything and everyone. The ever-increasing demands for digitalization, higher data rates, lower network latencies, better energy efficiency, reliable worldwide connectivity with the ultimate goal of automating much of daily life became real. The ITU recommendations has predicted that in order to satisfy the emerging requirements, network capabilities of serving more users with higher data rates need to overcome substantial limits on hardware and channel conditions [21]. It became apparent that more spectrum resources are required to be allocated for the next generation of wireless communication networks, as existing systems spectrums aren't sufficient and making it challenging to enhance performance within the limited available bandwidth [23].

Based on the system requirements, more advanced spectral efficiency techniques and other enabling solutions were expected, which ultimately led to explosion of numerous facilitating technologies, differentiating enablers for millimeter wave communications, hardware enhancements, wireless network functions virtualization, mobile cloud computing, things (device-to-device) communications, and several associated and supporting radio access techniques (e.g. bandwidth parts

(BWPs), bandwidth adaptation, orthogonal frequency domain multiple access (OFDMA) with flexible subcarrier spacing, digitalization of massive Multiple Input Multiple Output (MIMO) antennas, advanced beamforming techniques to direct signals precisely, improving coverage, capacity, and reducing interference, particularly in high-frequency bands) [24],[25].
The ultimate goal for the 5G mobile networks, to address a significant leap from previous generations in terms of required speed, connectivity, and functionality, has resulted in major paradigm shift not only through the invented technical solutions. To accommodate them, 5G technology development brought out a breakthrough via: New Radio (NR), new core and new architecture [19], [20].

Figure 1 depicts a high-level overview of the technology components for 5G, referencing to 3GPP TS23.501 [19], which unveils further detailed technical specification for the 5G system architecture and the functional descriptions of the particular Network Functions (NF), entities and interfaces.

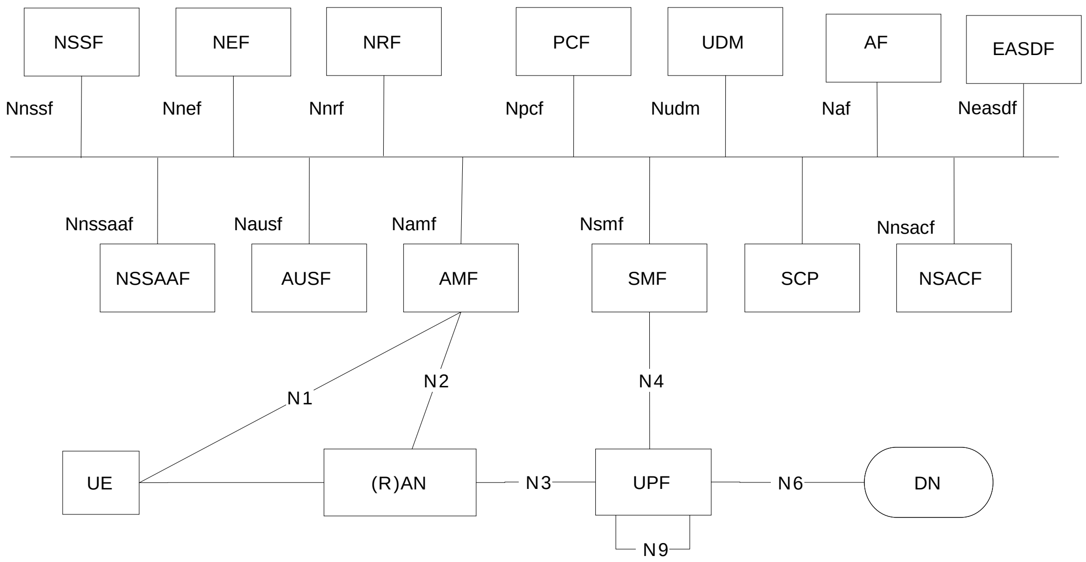


Figure 1: Overall 5G system architecture [19].

General concept for 5G system architecture has been to support data connectivity and services in a way that minimizes the need for interactions. For that reason, the scheme considers enabling deployments to use techniques such as e.g. Network Function Virtualization and Software Defined Networking, so that service-based interactions between involved Network Functions are leveraged. Entities can communicate directly, referred to as direct communication exist between the NF services in the Network Functions described by point-to-point reference point between any two network functions (e.g. AMF and SMF) [19]. Furthermore, some more detailed principles, which 5G architecture is compliant with, are to:

- Separate the User Plane (UP) functions from the Control Plane (CP) functions, allowing independent scalability, evolution and flexible deployments e.g. centralized location or distributed (remote) location.
- Enable each Network Function and its Network Function Services to interact with other NF and its Network Function Services directly or indirectly (the architecture enables interface bus system to help interact and route messages directly whenever possible).
- Minimize dependencies between the Access Network (AN) and the Core Network (CN).

5G radio access network (R)AN component plays a crucial role in the 5G system, as it is responsible for the wireless communication between user equipment (UE) and the core network (5G Core or 5GC). The NG-RAN architecture in 5G differs from the preceding radio access technologies in three main aspects: the logical split of the base station, control plane and user plane protocols separation and flexible deployment of its functions. Contrary to 4G RAN consisting of monolithic 4G base stations (eNodeBs), the NG-RAN is composed of radio nodes (gNBs) that can is split to different units: Central Unit (CU) and Distributed Unit (DU). A gNB may consist of a gNB-CU and one or more gNB-DU(s). As shown in Figure 2, this split architecture allows for the separation of control and user planes, distribution functions across different physical locations (e.g., the CU can be centralized for multiple DUs), enabling more efficient and low-latency operations.

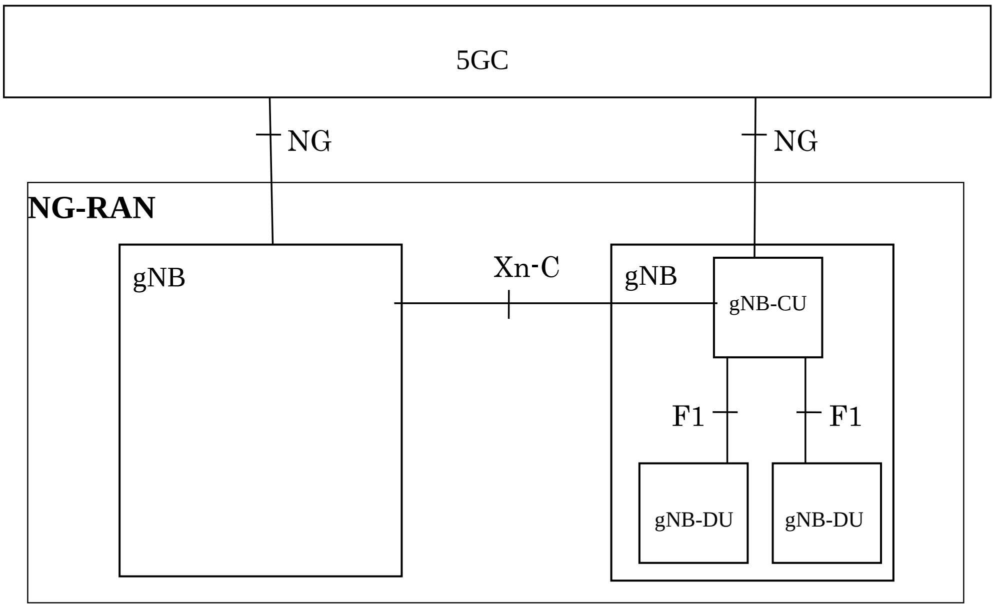


Figure 2: NG-RAN architecture [26].

Several implementation options for splitting the gNB are possible. 3GPP procedures adopt a split where the CU deploys the higher radio protocol layers from CP protocol stack and hosts (terminates) control plane from the air interface, while the DU deploys the lower layers operations and supports user plane maintenance. Detailed split of the layers corresponds to radio protocols architecture terminated between the gNB and the UE, as defined in 3GPP TS38.300 [20].

1.1.2 5G Air Interface

The air interface in 5G is the communication link between the User Equipment (UE) and the base station (gNodeB) (Figure 3). It plays a critical role in managing how data is transmitted over the radio waves, ensuring efficient and reliable communication. The 5G air interface, also known as New Radio (NR), includes various protocol layers and procedures designed to meet the stringent requirements of 5G. It is also organized in a split to user plane and control plane.
The user plane of NR radio interface consists of the physical layer (PHY), medium access control (MAC), Radio Link Control (RLC), Packet Data Convergence Protocol (PDCP) and Service Data Adaptation Protocol (SDAP), as illustrated in Figure 4.
The control plane of NR radio interface consists of the physical layer (PHY), medium access control (MAC), Radio Link Control (RLC), Packet Data Convergence Protocol (PDCP) and Radio Resource Control (RRC), as illustrated in Figure 5.

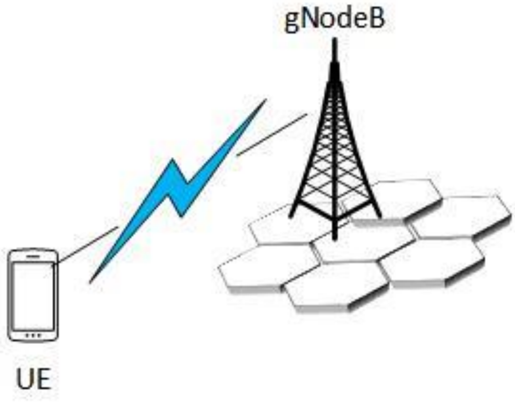


Figure 3: 5G air interface.

The radio interface protocol layers have different responsibilities and govern different functionalities [20]:

- PHY layer is responsible for the actual transmission and reception of data over the air interface. It handles the encoding, modulation, and demodulation of signals. This layer supports various modulation schemes, with adaptive modulation and coding to optimize data rates based on channel conditions. It ensures that the UE is synchronized with the gNB with proper timing alignment, essential for maintaining seamless connection.
- MAC layer controls access to the physical resources provided by the PHY layer, manages how data packets are transmitted and received over the air interface. It is responsible for scheduling and radio resources allocation, error correction by allowing the retransmission of data packets that were not correctly received with HARQ (Hybrid Automatic Repeat Request) method. MAC also maps logical channels to transport channels, determining how different types of data (e.g., control signals or user data) are transmitted and manages the prioritization of different traffic types.
- RLC layer is responsible for segmenting and reassembling data packets, ensuring reliable data transfer between the UE and the gNB, and data transfer modes assignments: Transparent Mode (TM), Unacknowledged Mode (UM), and Acknowledged Mode (AM)—each providing different levels of reliability and overhead.
- PDCP layer's role is to ensure efficient and reliable data transfer between the UE and the gNB, by header compression, data integrity and security provision and data duplication support (e.g. duplication of packets across multiple paths in multi-connectivity scenarios to improve reliability and reduce latency).
- SDAP layer is responsible for handling Quality of Service (QoS) flows, mapping them to data radio bearers, ensuring that data with different QoS requirements is treated appropriately and managing the flow control is maintained according to what QoS requirements target.
- RRC layer is critical for controlling the radio resources and managing the connection between the UE and the gNB. It configures each of the lower layers. This is the layer that handles connection setup and its maintenance, mobility management, including configuration of the measurements and resulting decisions from measurements results (e.g. handovers).

The Radio Resource Control (RRC) protocol that defines the signaling syntax for radio, procedures and principles performed in RRC layer is the key enabler for the air interface operations and making the communication link feasible. It is specified in 3GPP TS38.331 [27] and operates in 5G radio as the third layer (L3) of the radio interface protocol stack and is component of the control plane. Its functions are critical for all the operations in 5G, directly impacting the performance, security, and user experience of mobile communication services. The protocol procedures differ, depending on the state machine modelled for it, as illustrated in Figure 6.

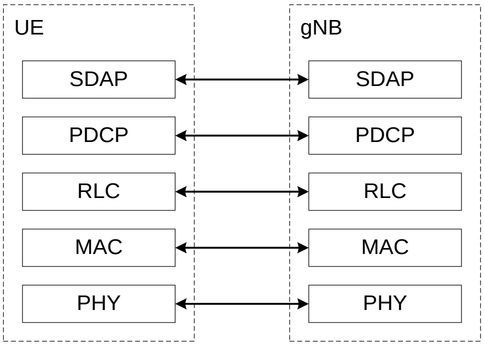


Figure 4: User Plane protocol stack in radio [20].

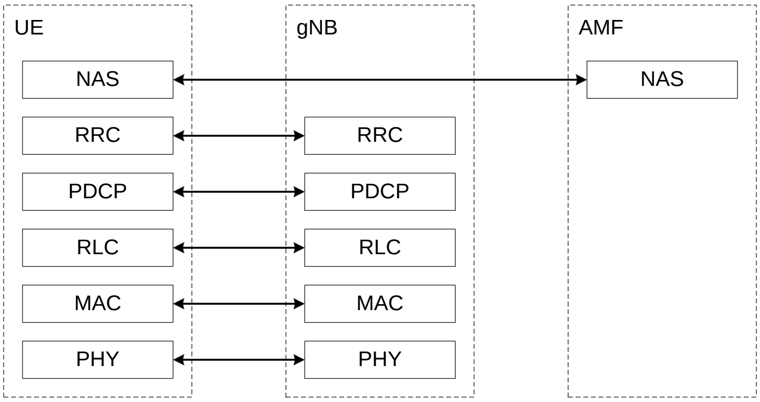


Figure 5: Control Plane protocol stack in radio [20].

A UE can be either in RRC_CONNECTED state or in RRC_INACTIVE state when an RRC connection has been established. If this is not the case, i.e. no RRC connection has been established (no RRC connection is active or suspended), the UE is in RRC_IDLE state. A UE has only one RRC state in NR at one time. By transitioning between states, the RRC protocol helps to manage and save UE's battery life as well as efficiently manage network resources. For example, in RRC_Idle, the UE saves power by not maintaining an active connection, but it can quickly transition to RRC_Connected when needed.

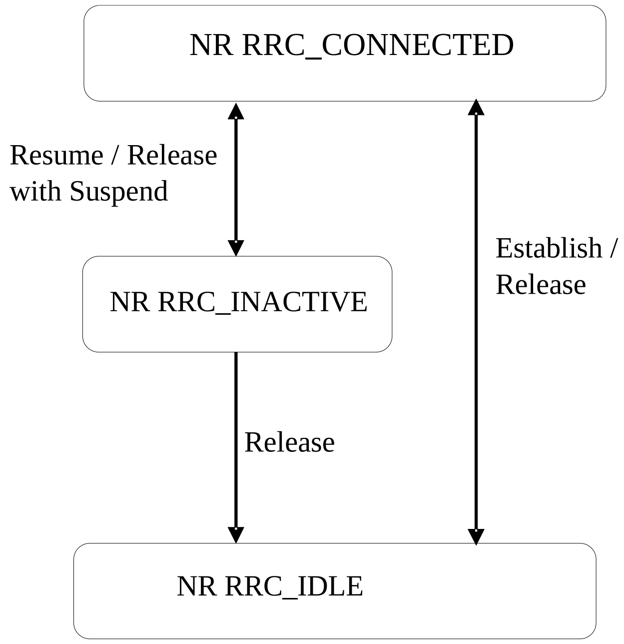


Figure 6**:** UE state machine and state transitions in NR [27].

The RRC states are characterized as follows:

- RRC_IDLE:
  - There is no active UE context on the network side, and UE controls mobility by itself, based on given in advance network configuration.
  - The UE monitors Short Messages transmitted on the lower layers, and monitors a Paging channel for CN paging using 5G-S-TMSI; performs neighbouring cell measurements and cell (re-)selection; acquires system information and can initiate RRC connection, performs logging of available and pre-configured measurements together with location and time for logged measurement configured UEs.
- RRC_INACTIVE:
  - There is UE context on the network side available temporarily and only in a given RAN - based notification area and the UE controls mobility by itself, based on given in advance network configuration.
  - The UE monitors Short Messages transmitted on the lower layers, performs neighbouring cell measurements and cell (re-)selection, performs RAN-based notification area updates periodically and when moving outside the configured RAN-based notification area, acquires system information, performs logging of available measurements together with location and time for logged measurement configured UEs; performs limited transmission of data (i.e.: small data) or reference signals for positioning, supports RRC connection resume.
- RRC_CONNECTED:
  - There is UE context on the network side available and actively maintained, there is major network control for all the operations, network-controlled mobility available and active.
  - The UE monitors the lower layers, performs neighbouring cell measurements and cell (re-)selection, acquires system information, monitors control channels associated with the shared data channel to determine if data is scheduled for it, provides channel quality and

feedback information, performs RRC Connection management (upon the connection has been established), ensuring that the connection remains stable, and adapting to changing radio conditions, performs measurements together with reporting or any other data transmission.

The RRC Connection management is initiated by RRC Connection Setup (Figure 7) and last until RRC Connection Release message is initiated (when the communication session ends and RRC protocol is responsible for releasing the connection, freeing up radio resources, Figure 8Figure **8**: RRC connection release [27].).

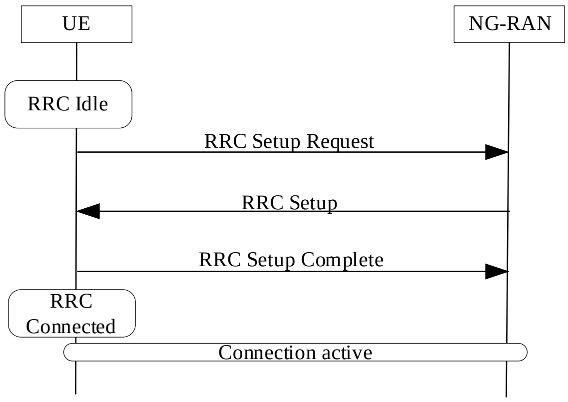


Figure 7: RRC connection establishment [27].

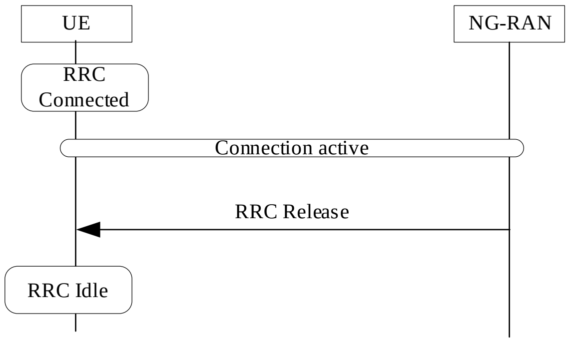


Figure 8**:** RRC connection release [27].

The sequence of procedures enables RRC state transitions. Once the UE transitions from RRC Idle state to RRC Connected it remains in the active UE connection with the network. Vast of the essential operations, which are critical for both: the UE active connection and network operation happen in RRC Connected state. In this mode, the RRC layer and protocol handle more detailed procedures, with some examples as follows:

- Security: The RRC protocol handles the establishment and management of security. It configures the security keys for encryption and integrity protection, ensuring that both user data and signalling messages are secure and the communication between the UE and the network is secure and protected from unauthorized access.
- Mobility Management: The RRC protocol manages UE’s mobility and resulting from it the handover process, which allows the UE to move from one cell to another or between different gNBs without dropping the connection. It ensures that the UE maintains a stable connection as it moves, by coordinating the transfer of the UE's connection context from the source to the target cell.
- Lower Layer Configuration: The RRC protocol configures all the lower layers (PHY, MAC, RLC, PDCP) with necessary parameters to facilitate resource allocation, power control, transmission settings, and transmission control through configuration of functionalities such as segmentation, error correction, and transmission mode.

- Measurements configuration: The RRC protocol configures the UE with network configuration for radio signal (frequencies) to be monitored, measurements to be performed and the way of reporting, which is essential to steer the UE select the most suitable cell to camp on based on signal strength and to steer the UE by the network criteria to operate in the network smoothly.

The contents of each RRC procedure is specified in [27] using ASN.1 to specify the message syntax and defining handling of the signaling in firmly established way, to make the communication between the UE and the gNB clearly defined for compatibility reasons. An exemplary signaling message encoding, according to [27], is given below:

```
-- ASN1START
-- TAG-MEASUREMENTREPORT-START

MeasurementReport ::=                SEQUENCE {
    criticalExtensions                   CHOICE {
        measurementReport                    MeasurementReport-IEs,
        criticalExtensionsFuture             SEQUENCE {}
    }
}

MeasurementReport-IEs ::=            SEQUENCE {
    measResults                          MeasResults,

    lateNonCriticalExtension                 OCTET STRING
OPTIONAL,
    nonCriticalExtension                     SEQUENCE{}
OPTIONAL
}

-- TAG-MEASUREMENTREPORT-STOP
-- ASN1STOP
```

This work will examine methods relying on the functions performed in Control Plane of 5G air interface, with particular interest of RRC layer operations in RRC Connected state. Due to its essential role and determining function on how 5G connection is ongoing, how measurements are performed or how data is transmitted and received between the UE and the gNB, the RRC protocol represents a significant importance for enabling any new application that require air interface signaling support.

1.1.3 Key 5G Traffic Use Cases

5G is designed to meet diverse requirements across various industries and verticals. The associated requirements of the supporting 5G system leads to broad variety of capabilities (Figure 9). Once some users need transmission of high-definition data while on the move, and the others create stationary congestion, there might be simultaneous need to support devices requiring very high data rate. The system targets latencies as low as 1 millisecond, especially for applications like autonomous driving and industrial automation. At the same time 5G offers high-speed data rates (up to 10 Gbps), supporting high-definition video streaming, Virtual or Augmented Reality (VR/AR), and other immersive applications. Yet, 5G promises to satisfy massive capacity to connect a massive number of devices (up to 1 million devices per square kilometer), with intention to supports IoT, smart cities, and dense urban areas.

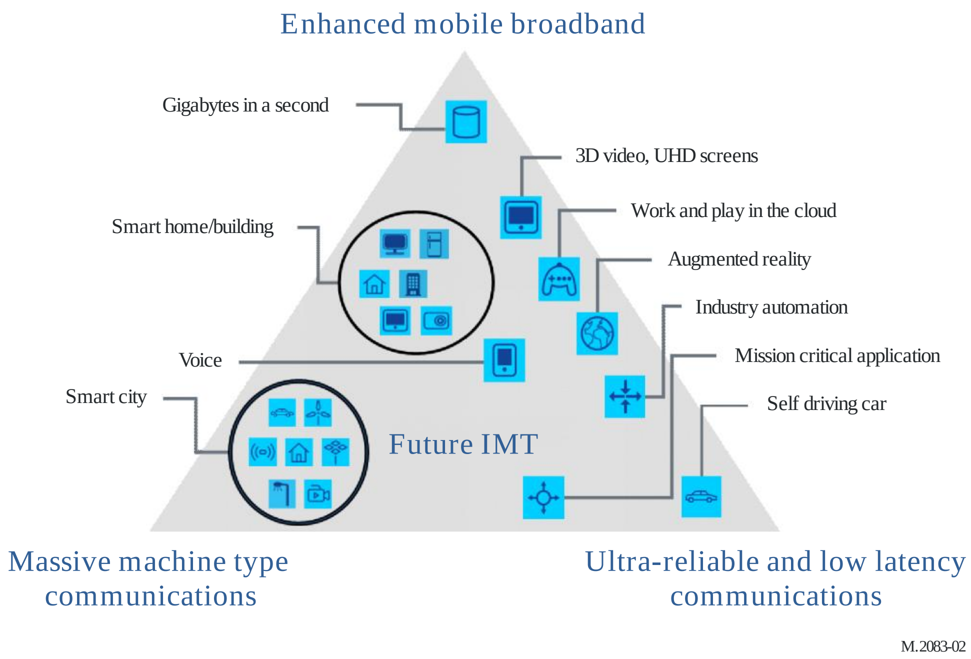


Figure 9: Usage scenarios for 5G and beyond [5].

Although, noting all the emerging trends, classification of the traffic types in 5G at the international level tends to streamline dimensions for the representation of 5G traffic. ITU envisioned three representative and nominal use cases for 5G networks: Enhanced Mobile Broadband (eMBB), Ultra-Reliable and Low-Latency Communications (URLLC) and Massive Machine Type Communications (mMTC) [5]. eMBB aims to address all the human-centric use cases for access to multi-media content, services requiring high data rates with improved performance and an increasingly seamless user experience. The services include video streaming, gaming or broad variety of mobile data generated at smart home building. URLLC has been classified to include wireless support for industrial automation or production processes, remote medical surgery, transportation safety, etc., all meeting stringent requirements for critical capabilities such as latency and availability. In general, this profile aids to support industry automation, mission critical applications or self-driving cars. mMTC is characterized by a very large number of connected devices typically transmitting a relatively low volume of non-delay-sensitive data. This profile covers smart city applications, as well as mission critical connections.

Devices connected through 5G are expected to exhibit different levels of important capabilities, including data rate, connection density, and latency (Figure 10: The importance of key capabilities in different 5G usage scenarios [5].). The fundamental technical performance standards and their corresponding metric definitions for the 5G radio interface are outlined in [5]. Additionally, the report in [21] aligns the three 5G application scenarios (eMBB, mMTC, and URLLC) with specific evaluation criteria. The standards establish a radio interface evaluation process that allows for a consistent technical assessment of overall performance, ensuring the achievement of the primary 5G goals The essential attributes of 5G can be assessed through simulation, analytical, or inspection-based methods for evaluating 5G performance. For instance, for the analytical evaluation, [50] establishes the specific indicators for 5G traffic relevant data gathering as defined in Table 1:Key performance indicators of 5G per traffic type.

Table 1:Key performance indicators of 5G per traffic type.

eMBB	URLLC	mMTC
- Peak data rate - User experience data rate	- Latency - Mobility interruption time	- Connection density

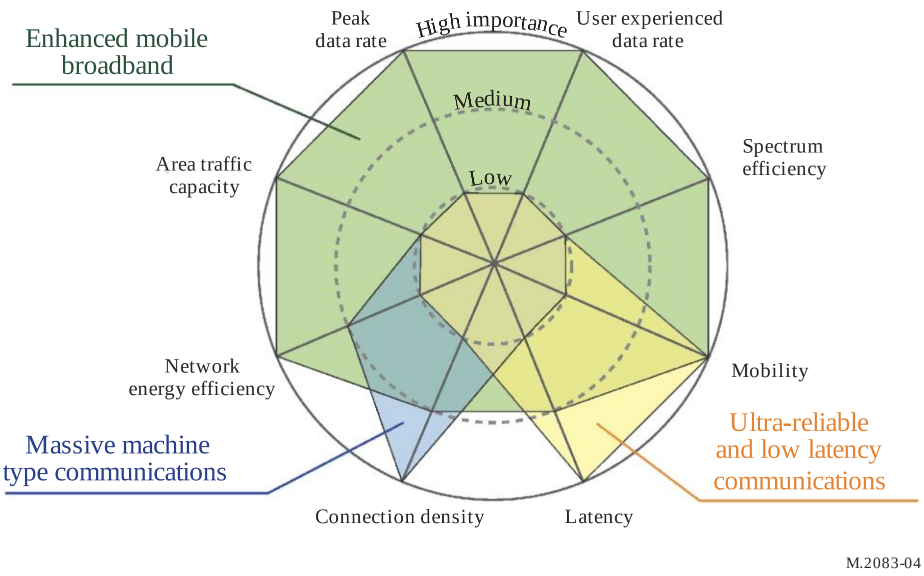


Figure 10: The importance of key capabilities in different 5G usage scenarios [5].

1.1.4 5G Network Performance Optimization Methods

In addition to key usage scenarios, architectural aspects, protocols and interfaces design, 5G standards include methods to facilitate and manage 5G deployment [9],[28],[29]. The techniques incorporate variety of network optimization solutions to ensure smooth and efficient operation.

One such method is the advanced Self-Organizing Network (SON), which targets the automation and efficiency of mobile network operations (Figure 11). It enables network management by automating operations such as self-commissioning (plug-and-play RAN nodes), self-healing (e.g., if a base station fails, SON can reroute traffic to neighboring cells, adjust power levels, or trigger backup systems, minimizing service disruption), and self-configuration. For example, SON function, running on the active network elements, can decide about adjustments, such as optimizing handovers, balancing loads between cells or tuning antennas' parameters. This automation allows for minimized operational activities spent on network setup, monitoring, parameter tuning, traffic and load management. Next Generation Mobile Network (NGMN) Alliance defined in [28] specific requirements covering expected for SON functionalities and interactions within the network, including network entities' locations, radio parameters importance, transport of the parameters across the involved network entities, and neighbor node data alignment need.

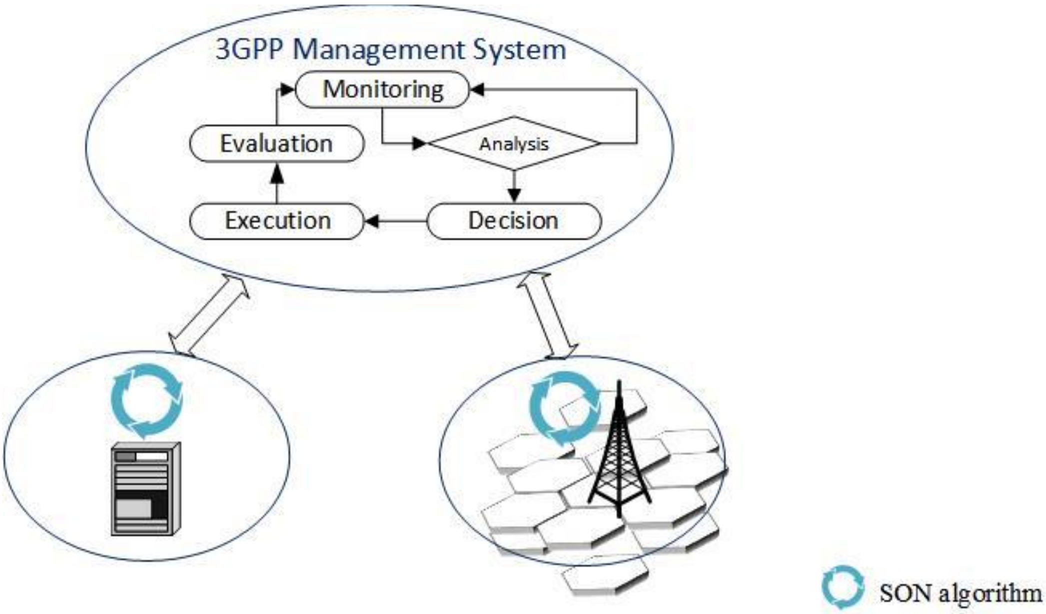


Figure 11: SON architecture [29].

The document emphasized that minimizing operational efforts through enhanced automation need compatibility in multi-vendor environments. The recommendations were completed by standardization of the SON architecture that accommodated the requirements in [29]. According to 3GPP TS28.313 [29], the technical solution relies on network entities that can run SON specific algorithms, in particular neighboring cells or cells clusters, as grouped by NG-RAN scope, which can be illustrated as in Figure 12. According to the SON architecture, optimization can be performed in a cell (and between relevant cells) managed by a SON server, which particularly considers the states of cells inside its management. Accordingly, a consideration is focused on how optimization performed by a network management system, which includes multiple radio stations and network operation management with SON servers is managing the radio stations in which network optimization is performed.

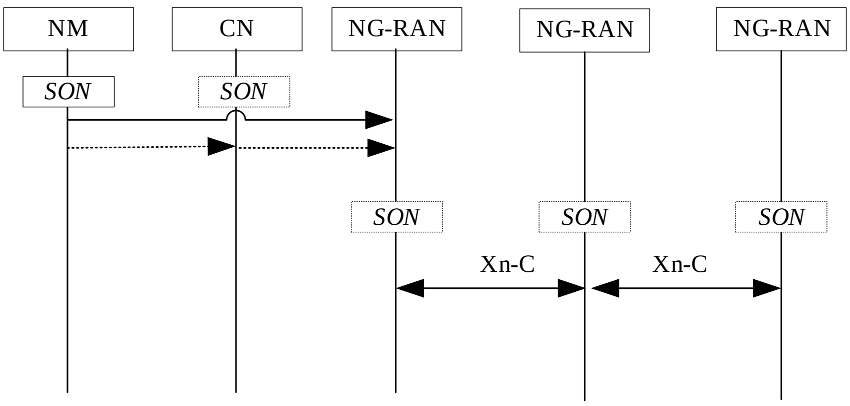


Figure 12: SON functionalities and interactions within the network.

Another method included in 5G is Minimization of Drive Tests (MDT), that smartly relies on the users' devices reports, to continuously gather data on network performance. The technique has been introduced by the need to optimize network performance monitoring by reducing the need for traditional, manual drive tests. Instead of relying on manual data collection, MDT relies on the UEs in the network, to automatically gather radio measurements during regular network usage (Figure 13).

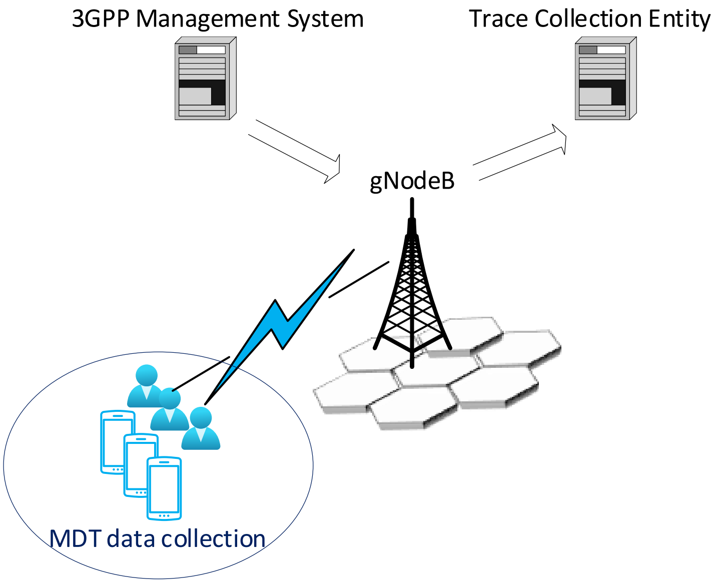


Figure 13: MDT architecture.

The data is recorded in a way that can be used by network operators to assess network performance in an offline manner through post-processing operations. To facilitate offline post-processing, devices generate detailed reports by tagging them with location information and timestamps. The enhanced network performance insights are promised through real-time data from actual users, and the ability to monitor network conditions across a wide area, including dense network deployments or indoor environments, where traditional drive testing would be unfeasible and costly.

3GPP defines technical details covering supporting 5G architecture [30] and data to be collected [9] in the MDT sessions, requiring downlink radio signal quantities measurement results for the serving cell as the key metric to be collected. The framework relies on Trace Functionality originating from Network Management system, where the MDT configuration is triggered. It can be provided to the radio directly or through the CN. In any case the configuration is finally proxied to the end user vie RRC message, through regular radio configuration, and determines what metrics are in the scope. Depending on the original configuration and metrics list within, the radio procedures for NR MDT differentiate:

- Immediate MDT, where the network can collect data with the UE involvement in RRC_Connected state. For example, the UE can report radio signal measurements to the RAN via periodical or event-triggered ways.
- Logged MDT, where the network sends logged measurement configuration to the UE in its connected mode, but it is validated and activated when the UE enters RRC Idle or RRC Inactive state. This configuration results in deferred reports (i.e., non-real time).

The UE reports the measurement results to the NG-RAN, often as a part of regular operation (e.g. for radio link monitoring). Specifically, MDT measurement results are collected from the air interface to the NG-RAN through RRC protocol signaling. It is possible to signal the data onwards to operators' Network Management entities for further post-processing. This is enabled by sending the user's reports onwards by the NG-RAN to Trace Collection Entity (TCE). However actual processing and anticipated action on the data is out of the standard procedures scope.

The framework is a facilitator for user's data collection and its reporting back to the network entities, but does not impose specific requirements on how and where to analyze the data (Figure 14). By substituting the need for traditional, manual and intensive drive tests, MDT not only lowers operational costs but also provides operators with real-time feedback from the users.

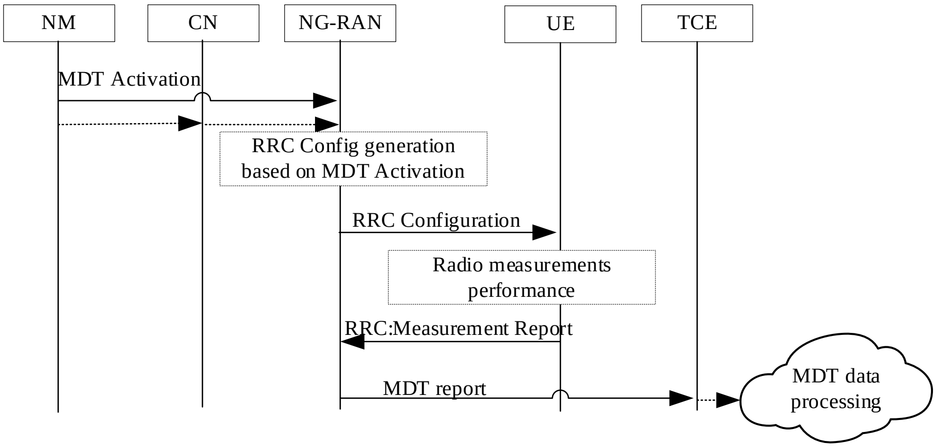


Figure 14: MDT framework.

The two network performance optimization mechanisms (SON and MDT) aim at automating huge efforts for the operators spent on collecting performance measurements, adjusting network configuration parameters, network planning and monitoring tools configurations, including multivendor scenarios. Depending on a scenario, SON ensures autonomous adaptation of the network by an optimization algorithm tailored to a specific need, detailed solution and requirement (e.g., the network may autonomously change setting of frequency or resources restrictions and preferences for the resource usage in the different cells based on interference level). MDT specifically facilitates automate collection of performance measurements, especially from the active users. Detailed purposes for which both solutions are dedicated are listed in Table 2.

Table 2: RAN Optimization Use Cases per method.

RAN Optimization Use Cases			
Toolkit	Task	Purpose	Key Data Producer
SON	Automatic Configuration: - Self-Configuration - Self-Commissioning - Self-Healing - Self-Optimization	- Mobility Robustness Optimization - Load Balancing/QoS Optimization - Energy Efficiency Optimization - Capacity and Coverage Optimization - RACH Optimization - PCI Configuration - RRM Resources Configuration	NG-RAN/ NR cell
MDT	Automatic Data Collection	- Mobility Optimization - Coverage Optimization - RACH Optimization - QoS Optimization	UE

In practice, SON is focused on streamlining network operations. SON-enabled optimization use cases are integral to the efficient management and operation of modern mobile network in high traffic and dense areas, where SON can adjust the distribution of network resources across cells to prevent congestion and maintain service quality (Load Balancing). Notably, SON-enablers become powerful in NR cells, operating in autonomous relation with neighbors, ensuring that users experience consistent service quality across different cells, even in challenging environments such as stadiums or transportation hubs (Coverage and Capacity Optimization, Mobility Robustness).

MDT, in practical deployments, has demonstrated its key value in collecting measurement results from the regular users in the field in a real time manner. The fundamental concept aimed at replacing dedicated and costly drive testing performed for network optimization. To facilitate the replacement, the feature involves commercial devices of cellular network users and makes usage of their data that are collected anyway (e.g. for mobility purposes) [9].

The two technologies complement each other, working together to optimize the network and ensure that users are satisfied with their service. Both functions support the following fundamental use cases:

- **Coverage optimization**, which consists of detecting and optimizing cell poor coverage and over-coverage collectively, coverage maps planning, cell coverage and capacity monitoring according to operator specific deployment strategy and requirements on traffic models, coverage holes detection.
- **Mobility Optimization**, which consists of mobility events observation (handovers), including detection of mobility events resulting errors causing radio link failure (RLF) due to unsuccessful handovers, detection and correction of errors in the mobility configuration, etc.
- **QoS Optimization**, which consists of user's quality of experience monitoring, verification whether quality of service experienced by the end user is in line with the performance target defined in the planning strategy, data rate and cell throughput observations, etc.

The overlapping optimization purposes approached by the two originally different methods reveal the unprecedented importance for coverage, mobility and QoS optimization, laying the groundwork for self-configuration, self-healing, and self-optimization based on automated data collection capabilities. The specified metrics as essential performance indicators for RAN Optimization-related use cases can be categorized as shown in Table 3: Key performance indicators per RAN Optimization use case..

Subsequent 3GPP standard releases are expanded these autonomous functionalities, incorporating more advanced solutions and enabling new dimensions and greater automation. These enhancements are particularly relevant in the context of 5G, where the complexity and scale of network operations necessitate advanced network managerial capabilities. In addition, these approaches resonate with industrial interest and hope in what AI can accomplish. It is envisioned that computational abilities of appropriate ML techniques can empower SON and MDT with intelligence and shape new applicability of the SON and MDT features in continuously developed 5G networks.

Table 3: Key performance indicators per RAN Optimization use case.

Coverage Optimization	Mobility Optimization	QoS Optimization
- RSRP - RSRQ - SINR	- RSRP - RSRQ - SINR - Interruption time	- RSRP - RSRQ - SINR - Latency - Throughput

1.2 AI/ML Philosophy and Development

1.2.1 Power of AI

Early Artificial Intelligence was a self-discipline. In the early 1950s, when the first use of the term Artificial Intelligence has been coined, single computers were taking up entire floors. Even with their enormous size, they had much less processing power than todays' smartphones. Despite, it was recognized the machines exhibit intelligent behavior, as enable mechanization process of human's thought or intelligence.

In the very beginning, computer programs were designed to generate outcomes to problems that could be presented as mathematical formulas and already at that time great potential was seen to solve computational tasks with such programmed intelligent machines. In addition, AI used approach called the Symbolic Systems [31]. This approach allowed machines to act in a way that seemed intelligent, but in reality it was just symbol matching without learning the context. The symbolic systems approach guided early programmers to create a database with known outcomes and react on a command from the set of created symbols. The machines, working on the system basis, could do things, but the problem with these systems was that the experts created long lists of matching patterns. The capabilities of AI programs were limited. Even the most impressive ones, could only handle trivial versions of the problems from todays' perspective. Even with these challenges, symbol systems were still the core of AI for years, yet in the end running into combinatorial explosion.

There are certainly processes that computers are very good at. In fact, there are many learning, computational and parallelized tasks where they're much better than humans. Human brain power has not changed much in thousands of years, whereas computers have improved from gigaFLOPs in the 1980s, teraFLOPs in the 1990s, petaFLOPs in 2008, and exaFLOPs in 2018 [2], according to [32]. A crude comparison of a leading supercomputer, a typical personal computer; and the human brain is presented in Table 4: Comparison of a supercomputer, personal computer; and the human brain processing capabilities, as per [32]. The growth from "giga" (billion) to "exa" (quintillion) represents a tremendous increase in scale, and is associated with extreme-scale computing capabilities that computers currently have. As technology evolves, innovations in semiconductor technology, distributed computing, and cloud infrastructure have fueled this growth too.

Table 4: Comparison of a supercomputer, personal computer; and the human brain processing capabilities, as per [32].

	Supercomputer	Personal Computer	Human Brain
Computational units	10^6 GPU + CPUs 10^{15} transistors	8 CPU cores 10^{10} transistors	10^6 columns 10^{11} neurons
Storage units	10^{16} bytes RAM 10^{17} bytes disk	10^{10} bytes RAM 10^{12} bytes disk	10^{11} neurons 10^{14} synapses
Cycle time Operations/sec	10^{-9} sec 10^{18}	10-^{9} sec 10^{10}	10^{-3} sec 10^{17}

Exponential increase in data generation and the need of handling data at the exabyte level allow to quite confidently predict that no sector will be left untouched by AI [33]. Serendipitously AI capabilities, effectively meets the emerging socio-technical evolution, thus are hailed as promising tool to bring remedy in automating much of life. The impact of AI on society will become integrated in the daily life together with digitalization trends and advancements in technology.

1.2.2 AI Categories

Since the first wave of symbolic AI, the domain has radically transformed. Various theories of "intelligence", and challenges towards AI were explored. The famous Chinese Room Argument, arguing that computer's programs cannot afford imagination and intellect as mankind, was the subject of debate for decades [31],[34],[35],[36]. Under the argument perspective, 'strong AI' was the ambition that suitably programmed computers (or the programs themselves) could understand the task they are programmed to and could have other mental capabilities like the humans. Such computers could really act intelligently by possessing projection ability to predict the possible meanings of a situation as seek, to select from the multiple meanings within a given situation, possibly novel responses to a situation. By contrast, 'weak AI' is the view that computers are unable to conceive or predict responses beyond selecting from multiple replies they were programmed to. As the programmed beforehand outcomes were precise rule-based procedures (known as 'algorithms') prepared by human experts, the remaining computer task was to follow 'intelligent' sequence of the procedure steps, to decide how to respond intelligently to a given situation [37]. This theory was

[2] 1 exaFLOP = 10^{18} floating point operations per second

claiming that no machine can think, truly understand or have cognitive states and the actual intelligence of machines is 'weak'.

In the light of scientific studies and technology possibilities, the field of AI evolved significantly. It is recognized that weak AI has made significant advancements [38]. Symbolic AI, which at its best is constrained to a limited data and environment does not change much over time, (the rules are strict, and the variables are quantifiable), has been subject to adoption to 'big data' driven approaches. Through developed electronic computer's capabilities, much more sophisticated algorithms have been developed. In the course of machine learning trends, the AI algorithms come up for learning process, which involves several key steps through which the algorithm improves its performance on a given task by analyzing data. In most advanced versions, the process involves collecting data, selecting and training an algorithm, evaluating its performance, and refining the algorithm through iterations to improve its predictive accuracy and generalizability. Through these steps, the AI algorithms are tuned with various updates, continuously achieving a broader understanding and meaningful engagement across a variety of different contexts [39]. In fact, this represents an effective learning approach and result in novel responses to a situation, instead of merely giving automatic codified answers.

Whereas the original split to weak and strong AI might appear to be dated and still subjective, even the 'weak' AI bypasses human efforts surpassingly. Industrial voices give credits to 'weak' AI, claiming it is probably the only way to efficiently exploit Big Data [33]. The practical and unquestionable value of the AI in processing Big Bata, despite its limitations in achieving true intelligence, becomes nowadays truly relevant and successfully applied across number of sectors.

1.2.3 Machine Learning

Machine learning (ML) refers to an automated method that utilize a set of rules on data. At its most basic, machine learning uses programmed functions that receive and analyze input data to predict output values within an acceptable range. The rules lead to a transformation of an input/data/network to an outcome, so that the transformation's outputs are considered as responses, remaining in relation to the inputs. ML techniques differentiate in the process different data collection approaches (e.g. grouped data) and different functions employed to derive outcomes out of the collected data (e.g. statistical models). The functions can be employed depending on the nature of the data, the desired outcome, and the complexity of the relationships within the data. Most advanced approaches facilitate further the process with a step/rule that achieves learning by making gradual improvements based on individual operations, or by applying evolutionary principles to the function to yield gradual improvements [37].

Big Data-driven approaches led to employment of numerous statistical methods or scientific computations to derive relationship between variables (depending on the nature of the data, the desired outcome, and the complexity of the relationships within the data). These statistical methods or mathematical computations incorporated into the comprehensive ML process, are often referred as a core algorithm (or model) determining the ML method. In fact, numerous ML algorithms account for statistical methods with some advancements. For instance, the ML process can incorporate a function that performs data series analysis and learns how to perform a task based on the available data using certain pattern recognition. The preprocessed data (e.g., clustered or labeled) can be fed into a function in the form of input-output pairs (e.g. for supervised learning), which can get expanded, if the function gets more data. The task can be to correlate input with output, by simple matching or computation and the statistical method can be used to model the relationship between two variables (input and output) with different combinations. The more data the function gets, the more it can correlate and automatically generate new sets of outcomes that it didn't have before. Yet, in advanced approaches as new data is fed to these functions, they learn and optimize their operations to improve performance, developing 'intelligence' over time. I.e., the function can make predictions (based on variety of instructions), and it can be tasked to assess an error between the predicted output and the actual output that becomes calculated or correlated. Further, there can be internal parameters in the function (weights) that are monitored and adjusted to minimize the prediction error, which is

typically done using an optimization technique like gradient descent. The ML process incorporating such self-learning component and statistical model is often referred as model-based.
An exemplary comparative table between traditional approach, where analytical models are built from historical data only (historical regression model), and ML process with advancements on self-learning is shown in Table 5: A comparison of traditional and ML approach, adopted from [33].

Table 5: A comparison of traditional and ML approach, adopted from [33].

	Traditional (static approach, not/not well connected to a system)	Machine Learning (dynamic approach, interconnected with system/environment)
Prerequisites	A sample of the data that is representative of what is tried to be predicted	System must be connected to its environment (to interact with)
Model Objectives	Objectives are determined out of the sample data	Objectives are determined
Model Training	The model is initialized from the past data	The system is trained with a large volume of data, can be initialized from real time or past data
Model Updating	Requires reset to update the model	The system is self-learning, it will adapt between action and reaction in a closed loop principle.

ML accounts for numerous options in terms of dataset classification and processing. Decision on which approach, belonging to comprehensive ML methodology, needs to be adapted for a given scenario, depends on data nature, environment and desired purpose.

1.2.4 Machine Learning Categories

ML, which enables applications to operate in an "intelligent" way, is a fundamental component of AI and is essential for uncovering insights from data. There are several kinds of ML, including supervised, unsupervised, semi-supervised, and reinforcement learning [40]. The types of data processed in ML can range from unstructured and semi-structured to well-structured and metadata. The effectiveness of different ML techniques varies across domains based on their underlying principles and objectives (Figure 15).
Unsupervised learning focuses on analyzing unlabeled data sets and can work with diverse input data, aiming to uncover significant trends and patterns. It necessitates large volumes of data to detect hidden relationships, group inputs, and form various categories. In contrast, the supervised learning approach can operate effectively with smaller data sets to estimate necessary parameters and quickly produce desired outcomes, such as clustering or classifying data. This method typically involves mapping inputs to outputs by demonstrating the relationships between different variables and known results, relying on labeled data and guided supervision during the learning process. It is particularly suited for addressing straightforward problems.

On the other hand, reinforcement learning evaluates optimal behaviors in specific contexts or environments by analyzing relevant data through software-enabled agents that use a reward/penalty system based on observations. This learning type aims to leverage insights from the environment to make decisions that enhance rewards or reduce risks and failures. Through this approach, an algorithm learns by engaging with its surroundings, receiving feedback in the form of rewards or penalties, and striving to maximize total rewards over time. These methods are recognized as powerful tools for training AI models that promote automation, improve the operational efficiency of complex systems, and tackle various real-world challenges across multiple fields [41].

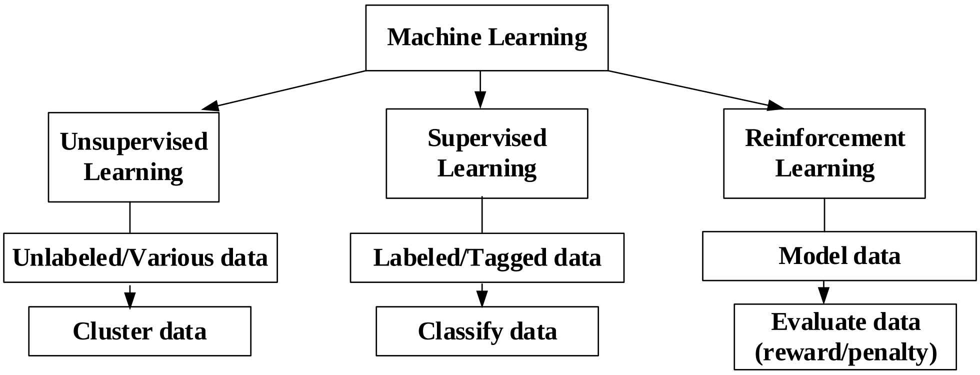


Figure 15: Different machine learning approaches for data classification [41].

The application of ML often instantiates hybrid methodologies. Right approach selection, including appropriate identifying and grouping related data is a key to build effective ML-based solution, as the outcome of a ML method can be more accurate and meaningful.

1.2.5 Artificial Neural Networks and Deep Learning

Artificial Neural Networks (ANN) are advanced methods for data analysis that utilize ML-based approaches and their combinations. They denote statistical models that emulate the function of a network of human neurons. Inspired by the human's brain functionality, ANN organize data into inputs that are translated into signals which are passed through a network of artificial neurons to generate outputs that are interpreted as responses to the inputs [37]. The models based on a collection of connected units or nodes called artificial neurons are combining the biological principle with advanced statistics methods to derive an outcome. Adding more neurons and layers allow ANNs to tackle more complex problems.

Neural network models are viewed as a functions solving problems in domains such as pattern recognition, but with exceptional advancements by a parallel processing ability. They can perform more than one task at the same time.

ANNs with several layers refers to Deep Learning methodology [41]. It is a part of ANN-based machine learning approaches, which provides a computational architecture by combining several processing layers: input, hidden, and output layers with connected neurons, to learn from data (Figure 16). The main advantage of deep learning over basic machine learning methods is its better performance in learning from large datasets [41].

In the more detail subject of deep learning, there are several variants of ANNs. Convolution Neural Networks (CNNs) are multi-layered ANNs enabling convolution operation, responsible for detecting the most important features. Another class of ANN in a Graph Neural Network (GNN) that enables processing data to graphs representation. These approaches can implement different flavors of visualized representation of the observed data.

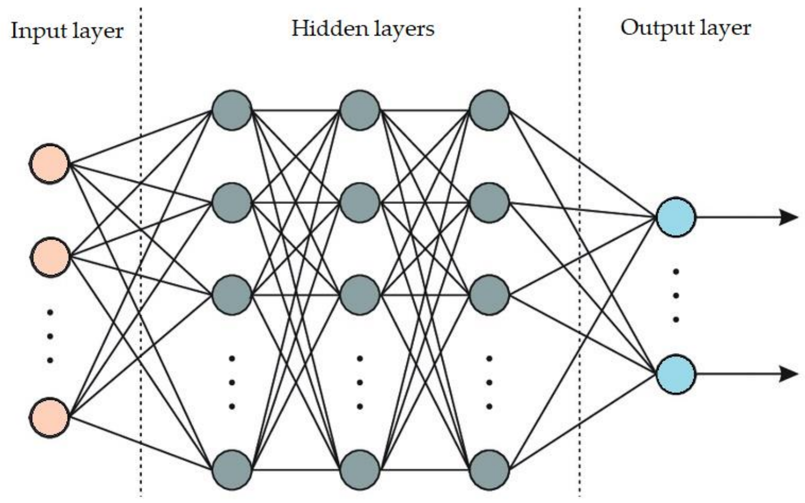


Figure 16: Model of Artificial Neural Network.

1.2.6 Machine Learning Core Algorithms Overview

Major activity of ML method is to build statistical model that hopefully reflect the important aspects of the object of study [42]. Common algorithms adopted to ML for that purpose include: Linear Regression Model, Decision Trees, Support Vector Machines, Neural Networks and more [41].

Linear Regression is a fundamental statistical method used to model the relationship between a dependent variable and one or more independent variables. It assumes that this relationship can be expressed in a linear equation. The simplest models of linear regression detect a relationship between two variables using a straight line. It is claimed to be not suitable for complex relationships where non-linearity comes to play [42].

Decision Tree is a method that models data using a tree-like structure of decisions based on the features of the data. The tree is composed of nodes (which test the value of a feature), branches (which correspond to the outcomes of the test), and leaves (which represent the output or decision). By sorting down the tree from the root to some leaf nodes, the method enables to classifies the instances. Instances are classified by checking the attribute defined by that node, starting at the root node of the tree, and then moving down the tree branch corresponding to the attribute value [41]. There is no assumption made in the method about the relationship between variables, inducing there can be non-linear relationships. It is suitable for data where the relationship between input variables and output is non-linear or involves complex interactions.

Support Vector Machine (SVM) is a method that high- or infinite-dimensional space aims to find the hyperplane that best separates the classes in the feature space. It constructs a hyper-plane or set of hyper-planes. Intuitively, the hyper-plane, which has the greatest distance from the nearest training data points in any class, achieves a strong separation since, in general, the greater the margin, the lower the classifier's generalization error. It is effective in high-dimensional spaces and can behave differently based on different mathematical functions known as the kernel [41]. SVM can efficiently handle non-linear classification by transforming the original features into a higher-dimensional space using kernel functions. The method can work with both linear and non-linear relationships but appears to be more complex than linear regression and decision trees, particularly when using non-linear kernels. It shows benefits for binary classification problems, especially where classes are not linearly separable. However, once applied to large datasets with many features, can be inefficient, as training can be computationally intensive.

Table 6 presents key differences and applicability of the methods:

Table 6: Comparison of Linear Regression, Decision Tree, and Support Vector Machine Methods.

	Assumption/Prerequisite	Applicability/Best For
Linear Regression	Linear relationship exists among variables, Continuous outcome needed (given the prerequisite holds)	Simple method, straightforward interpretations
Decision Trees	Versatile, Any kind of relationship, including non-linear or complex relationship	Flexibility and transparency
Support Vector Machines	Data is not linearly separable, and the relationships are complex	Adaptability, particularly in classification scenarios with non-linear boundaries

Each method has its strengths and choosing the appropriate method for ML algorithm depends on the specific characteristics of the data and the objective. The selection of appropriate ML algorithms is crucial as different algorithms can yield varying outcomes based on data characteristics.

2 Research goal

2.1 Research problem

The primary problem arises from:
- **Diversified optimization targets and 5G requirements** tightly coupled with the different 5G use cases: Telecom operators offering their subscribers 5G with 5G flexibility and diversification need to provide a network that copes with ambitious goal to accommodate varied demands. Considering only the three representative use cases: eMBB, URLLC or eMTC, which are introduced to be virtual slices, the supporting 5G system needs to offer broad variety of capabilities. It needs to satisfy, at the same time, often contradicting requirements. For instance, eMBB users demanding high data volume may co-occur with URLLC devices, which require ultra-low delays for small data transmissions. Although, the different 5G traffic types might be seen as virtual sub-networks (network instances), they may be created within the substantially same infrastructure, sharing the same physical platform and sharing the same radio interface. To be provided concurrently across the same RAN (or cell), network needs to run services that have different requirements on latency, reliability, throughput or mobility.

The paradigm nails further down to the implications with:
- **Deficiencies of existing automated solutions:** Network vendors, supplying telecom operators implement solutions for helping to facilitate 5G deployment, inheriting various optimization techniques practiced already in the previous mobile networks' generations: SON and MDT. These optimization toolkits help in automating huge efforts for the operators, but 5G maintenance appears much more complex due to the 5G traffic variation, flexible network deployments, as well as varying implementations and capabilities. Solutions for RAN optimization are scattered towards very specific purposes (e.g. load balancing or QoS). These limitations with effectiveness and efficiency of the known autonomous solutions is acknowledged by market insights. It is claimed that advances in 5G RAN have introduced new operational demands and have increased the complexity of managing and automating RAN operations. As such, existing RAN automation solutions no longer fit for purpose [11].
- **Integrating AI to network optimization solutions:** Intelligent networking capabilities that can optimize RAN operations and enhance automation are a recognized need. AI explosion, concurrent with mobile broadband advancements, emerges to be promising and suitable tool.

MNO industry concentration has taken on increased urgency to understand AI applicability to 5G. This focus is at initial phase on building requirements and frameworks at a generic level. The general idea of using AI/ML as components of network optimization is adopting an intelligence to machines. However, understanding positive trade-offs between additional complexity due to large AI workloads or cost vs. real benefits isn't fully explored. Implementation seeks to understand factors that determine actual feasibility, implications and gains. Pure requirements to apply AI to the automated optimization solutions cannot anticipate the advancements are affordable or guarantee successful uptake. The efforts lack focus on methodology how to close the persisted data types' division. In particular, how to integrate ML in context of the RAN optimization for main 5G use cases. Thus, further automatization enhancements by integrating AI into existing architecture and RAN optimization methodologies are in the risk to remain very specific and scattered, as the optimization methods are.
- **ML Algorithms irresolution:** Following the recognized limitations and the ubiquitous AI, there are associated developments in standardization bodies to enhance RAN optimization methods. The investigations recommend a framework, which guide further development of technical solutions. This is a milestone, however ML algorithms are assumed to be up to implementation [44]. Although scientific literature elaborates on ML algorithms applicability to telecommunication, there is still a gap between the optimization methods that are on the research agenda and the optimization methods that can handle real implementation. Lack of methodology how to identify ML algorithm itself and its inputs in context of the main co-existing 5G traffic use cases, leaves the effect of any automated RAN optimization method enhancement with AI ambiguous.

Overall, existing and deployed in 5G network optimization methods face several emerging challenges. These challenges are driven by the growing diversity of use cases and the wide range of 5G capabilities. The general idea of using AI/ML as a tool to adopt an intelligence to already existing automated solutions presents new challenges due to uncertainty on how to apply beneficial AI approach to diverse traffics within the network which currently already support various optimization methods. Lack of practical methodology on how to identify AI/ML algorithms inputs (from radio measures in RAN) leaves the uptake of AI very uncertain in terms of feasibility in real deployments. The AI can help operational disadvantages only if the three scattered dimensions: different traffic types, different RAN optimization methods and broad scope of ML-algorithms can be integrated together with a common ground.

2.2 Research questions

There are two RAN optimization automation solutions: SON and MDT utilized in practice in 5G RAN implementations, laying the groundwork for self-configuration, self-healing, self-optimization and automated data collection capabilities. MDT-enabled data collection is integral part of todays' networks operations due to substantial human's effort reduction. The tool has made data collection automation from any active connection feasible through collection of users' radio metrics and other data, via radio protocols. The radio metrics from users in a cell can get acquired, but actual optimization targets (Coverage, Mobility Optimization or QoS Verifications) based on the results are left to post-processing operations and vendor specific resolutions. On the other side, the automation being the key process on SON, is assumed to be facilitated by sophisticated algorithms, however the algorithms themselves remain so far implementation specific.

Individually the two features reveal great benefits: MDT automated data collection in RAN, SON having (black-box) algorithms-driven decisions located in RAN. None of the method is predictive, and cannot anticipate growing difficulty with classification problems. As the complexity scales, further broadening of automated solutions with new techniques in RAN may become counter-productive and add additional complexity. Especially in radio, which relies on very scare and valuable resource. To become easily adaptable, AI should be looking at the existing traffic. Machine learning algorithms seem suitable to be applied, if built on the greatest benefits of the solutions already adopted to live network operations. Therefore, the key questions raised for this research direction include:

- Is it possible to integrate AI into 5G RAN architecture with leveraged complexity, without imposing additional overhead?
- How contrasting requirements resulting from diversity of 5G usage scenario can get accommodated into an ML algorithm?
- What are the representative radio link measures that can be considered as ML algorithm input?
- Is there an ML algorithm that could be uniform baseline, so that AI can be applied in RAN in an affordable and sustainable way?
- Can AI reinforce existing automated methods in mobile network optimization to close the gap with traffic types' division?

2.3 Objective and strategy

Research objective for this dissertation, is to investigate practical ML-based method to support 5G network's traffic monitoring and optimization. For this purpose, the strategy is to:

- scope potential solutions by literature review,
- analyze 5G traffic characteristic,
- inspect real data generated in 5G network based on 5G drive tests and smartphones records,
- examine the datasets with statistical methods with the goal to identify data importance, relations and few dimensions for ML algorithm input
- process the datasets with an ML algorithm, tailored to exploration of the collected data and assessment of the applied ML algorithms performance,
- exploit approach how the method can get integrated to the existing RAN optimization methods,

- promote the method to standardization bodies, and to RAN implementation landscape.

2.4 Theses

In the context of the given research problem, it is hypothesized that there are some parameters in radio characteristics that are correlated. Identifying essential one and excluding redundant ones can limit the key parameters observation and thus ML algorithm input for robust AI utilization. In detail the theses split to:

1. **Thesis 1: Within radio characteristics set, RSRP is the metric with most important relevance for RAN monitoring:** Within a broader set of radio characteristics that can be associated with different 5G traffic types, a differential subset of key performance indicators matter. Within the key radio characteristics, RSRP plays a crucial role in radio link assessment and there is correlation of the other parameters with the critical determinant. It holds significant relevance for addressing the core aspects of network performance. This critical metric facilitates meaningful analysis. Chapter 4.3 addresses and validates the thesis.
2. **Thesis 2: Within ML algorithms set, Decision Tree proves to be the most suitable method for RAN monitoring in the context of performance prediction**: The applied algorithm needs to be capable of handling advanced regression with non-linear relationships among data. Data monitoring methodology with such ML algorithm, when applied to data from 5G RAN, proves efficient and suitable means for monitoring radio interface with potential to reduce transferred data volumes. Chapter 4.4 addresses and validates the thesis.

The theses will guide a focused approach to data selection and analysis, aiming to optimize both the relevance of data used and the effectiveness of the analytical method applied.

3 Literature overview

This chapter provides an overview of literature that became relevant to the underlying themes of this dissertation, helped to proceed with the questions that arose during the conducted research, and brought the research towards inventing new methodology. In particular, based on the studied material, this chapter provides a comprehensive overview of the role and impact of key radio metrics like RSRP, RSRQ, SINR performed and provided by end users to network monitoring and how various optimization methodologies approached the metrics utilization since 4G. It explores how the KPIs have evolved over time till 5G, and what is their relevance and application in the operating and functioning network. Similarly, the chapter reviews standardization and recent research trends and in context of development and importance of ML techniques in mobile networks and their impact on network optimization methods. The referenced studies review is organized in the following order:

- Sub-clause 3.1 explores on Network Performance Monitoring, addressing:
    - Evolution of optimization methods and drive test methodology in 3GPP standard, with differentiation of key performance indicators scope, split to UE-based and network-based, and radio metrics definitions evolution.
    - How exploration of the network monitoring and optimization methods is approached in research articles. The selected literature focus is to understand theories on radio interface observability and utilization of radio metrics, through various simulation and environmental scenarios of experiments (manual drive tests, through MATLAB-build system simulators, to simulated network conditions in indoor testbeds).
- Sub-clause 3.2 explores on applying ML methods to wireless telecommunication, addressing:
    - 3GPP developments on understanding ML methodologies applicability in 5G networks.
    - Key trends on how the topic is explored in research.

3.1 Network Performance Monitoring

3.1.1 3GPP standard developments

Key Performance Indicators (KPI) provide information that support effective network design and monitoring. No network can be successful without its surveillance and performance maintenance. The subsequent generations of mobile network systems demonstrate increasing complexity and capabilities. Together, Key Performance Indicators considered for network monitoring have evolved. Since the beginning of cellular telecommunication (2G/GSM era), radio perception of the radio frequency (RF) used for transmission has been critical and unavoidable metric of interest. In 2G, two critical metrics were: Received Signal Strength Indication (RSSI) for signal power and Downlink carrier-to-interference ratio (C/I). Manual drive tests were the primary method for collecting data about signal strength, quality, and coverage, with the main goal to identify areas with weak signals (coverage holes), and assess undertaken calls quality. The tests helped engineers understand how the network performed in real-world conditions. Early drive tests used basic equipment like portable signal strength meters, which displayed signal levels as engineers drove through different areas. The data collected was often recorded manually. All the collected data and the key metrics including signal strength, and basic call performance metrics such as call setup success rate and call drop rate were collected with specialized equipment. Although, the main struggle and cost has been the specialized equipment, network planning and optimization with these human's efforts were crucial for ensuring adequate coverage and quality of service.

With the advent of 3G networks (UMTS/WCDMA), the complexity of drive tests increased due to the introduction of data services, higher speeds (at least 144kbit/s), and more advanced network features. Besides signal strength and call quality, data collection necessitated for network observability expanded to include more advanced parameters like energy per chip over interference (Ec/Io), Common Pilot Channel Power (CPICH), and data throughput rates [46]. Consequently, the manual drive tests became more sophisticated with the use of laptops equipped with specialized software, GPS systems, and mobile phones or modems equipped to specialized cars, that were driving through the network connected to the network to monitor its performance. According to [47], testing scenarios began to multiply and included variety of reasons driven by operators' needs:

- Deployment of new base stations: as the very initial phase of drive test that is performed for a particular cell. Manual drive tests were performed before and after service activation of the new cell. In practice, it has been a scenario in which radio measurements were taken in a 'trial' mode for a new base station or cell to transmit radio waves. I.e. the radio transmission was tried in a "test mode" (i.e. cell barred for normal access), and initial collection of downlink coverage measurements of the new cell and neighbour cells was scheduled in the intended area of coverage improvement. This exercise was enabling initial area tuning (e.g. selection of an appropriate antenna for the new cell, adjustment of antenna tilting of the new cell and neighbour cells, etc.). Service of the new cell was enabled and physically started after such successful testing and initial tuning. With the note that deployment of new base stations is a continuous and long-lasting process for a new system, drive tests had to be continued to collect continuously the coverage measurements in the intended area to make sure appropriate network coverage is ensured.
- Land development and road infrastructure changes, such as construction of new highways or major buildings: the additional occasion for drive test has been reasoned by "an event driven manner", in cases areas where new highways or buildings are constructed and thus become potential areas which residing population will increase, or path loss will change. Operators usually performed drive tests in these areas of interest to monitor whether downlink coverage gets impacted or needs to be improved. If network performance improvement has been deemed necessary based on measurement results, operators were to take action by deploying new cells, adjusting antenna tilting of existing cells, etc.
- Customer's complaints: this has been given as another case of "an event driven trigger" for manual drive test, based on customer's voice, indicating downlink coverage concerns (e.g. coverage loses at their home, office, etc.). Under such circumstances, operators performed "drive tests" in the concerned area to observe signal quality and find a root cause. Drive tests to collect downlink coverage measurements were performed again after any corrective actions to check whether coverage has improved in the area to an adequate level.
- Regular periodic drive tests: As indicated in [47], operators were performing drive tests around any area being subject to substantial changes in a timely manner. However, due to frequent changes under regular land development, especially for smaller scale constructions, drive test were practiced by operators periodically, in order to keep up-to-dated understanding of the downlink coverage levels provided in their networks.

The introduction of 4G (LTE), that brought a significant increase in high-speed data services (speeds of up to 150 Mbit/s download and 50 Mbit/s upload), led to introduction even further, more detailed performance measurements, including [48]:

- Signal Strength (RSRP), indicating average level of the detected network's cell reference signal received by the UE.
- Signal Quality:
  - Reference Signal Received Quality (RSRQ): measurement for the quality of the cell reference signal received by the UE.
  - Signal-to-Interference-plus-Noise Ratio (SINR): measurement by the UE, for the quality of the signal by comparing the signal strength to the level of interference and noise.
- Throughput (DL/UL), referring to the data rate achieved by the user in both downlink (DL) and uplink (UL).
- Latency, defining the time it takes for a data packet to travel from the UE to the network and back, whereas lower latency is critical for real-time applications.

- Packet Loss, defining the percentage of packets that are lost during transmission. High packet loss negatively impacts the user experience, especially in for IP traffic assessments and video streaming.
- Handover related measurements, defining success rate of UEs switching from one cell to another without dropping the connection, crucial for maintaining service continuity.
- Radio Resource Control (RRC) Connections related measurements, consisting of connection setup success rate of establishing an RRC connection between the UE and the network, which is essential for initiating communication.

As of 4G, the measurements started to reflect the importance of the entire service delivery chain, from the core network to the user device and service availability. Due to the magnitude of the additional KPIs, new dimensions were developed as the measurements family related to network internal and network-dependant operations (so called network performance measurements and network counters) [49].

Starting from there, drive tests became even more critical for ensuring network quality from radio dimension, particularly in urban areas where network density and interference levels are high. Though, at the same time drive tests now had to cover multiple technologies simultaneously (e.g., 2G, 3G, and 4G), requiring more advanced tools capable of handling diverse data streams. The process began to be so overwhelming that triggered worldwide efforts in 3GPP to look for more automation with tools that could conduct tests autonomously, reduce human error, and allow for more extensive and frequent testing [50]. 4G noted the fundamental evolution and enhancements of drive test through the research and industry focus on selection of the key performance indicators that should be collected from UEs within the framework of drive test minimisation (MDT).

The MDT framework noted RSRP, RSRQ and SINR importance as radio KPIs (denoted as M1), that should be autonomously collected by devices to replace the rigorous drive tests [9].

While KPIs related to end-to-end network performance like throughput, latency, and other network performance counters are recognized under the framework as important, these are network-side measurements. The radio characteristics from the end users' devices are necessitated to be known to the network for determining the network conditions and users' perception. 5G hasn't changed the nominal scale of the key radio performance indicators, but adopted MDT framework for 5G equivalent definitions (as reflected in Table 7) for synchronization signals (SS) and channel state information (CSI).

Table 7: Detailed definition of key Radio performance indicators, as per [51],[52].

Reference signal received power (RSRP)	
Definition	Reference signal received power (RSRP), is defined as the linear average over the power contributions (in [W]) of the resource elements that carry signals of the interest within the considered measurement frequency bandwidth and configured discovery signal occasions (4G) or window duration (5G). (cell-specific reference) For RSRP determination in 4G: the cell-specific reference signals are used according to physical channels and modulation defined in [53]. For RSRP determination in 5G: - Secondary Synchronization signals are used according to physical channels and modulation defined in [54] to determine SS-RSRP. - CSI reference signals transmitted on a given antenna port 3000 are used, according to [53] to determine CSI-RSRP The reference point for the RSRP is the antenna connector of the UE. For 4G: if receiver diversity is in use by the UE, the reported value shall not be lower than the corresponding RSRP of any of the individual diversity branches. For 5G: - SS-RSRP is measured based on the combined signal from antenna elements corresponding to a given receiver branch. For frequency range 1 and 2, if receiver diversity is in use by the UE, the reported SS-RSRP value shall not be lower than the corresponding SS-RSRP of any of the individual receiver branches. - CSI-RSRP is measured based on the antenna connector reference point. For frequency range 2, shall be measured based on the combined signal from antenna elements corresponding to a given receiver branch. For frequency range 1 and 2, if receiver diversity is in use by the UE, the reported CSI-RSRP value shall not be lower than the corresponding CSI-RSRP of any of the individual receiver branches.

Reference Signal Received Quality (RSRQ)	
Definition	Reference Signal Received Quality (RSRQ) is defined as the ratio: - For 4G: N×RSRP/(E-UTRA carrier RSSI) - For 5G: N×SS-RSRP/ NR carrier RSSI to determine SS-RSRP - or N×CSI-RSRP/NR carrier RSSI to determine CSI-RSRP (where N is the number of resource blocks in the given carrier RSSI measurement bandwidth. The measurements in the numerator and denominator shall be made over the same set of resource blocks. A Carrier Received Signal Strength Indicator (RSSI), comprises the linear average of the total received power (in [W]) observed only in certain OFDM symbols of measurement time resources, in the measurement bandwidth, over N number of resource blocks by the UE from all sources, including co-channel serving and non-serving cells, adjacent channel interference, thermal noise etc. The reference point for the RSRQ is the antenna connector of the UE. If receiver diversity is in use by the UE, the reported value shall not be lower than the corresponding RSRQ of any of the individual diversity branches. For 5G, for frequency range 1 and 2, if receiver diversity is in use by the UE, the reported SS-RSRQ value shall not be lower than the corresponding SS-RSRQ of any of the individual receiver branches.

Reference signal-signal to noise and interference ratio (RS-SINR)	
Definition	Reference signal-signal to noise and interference ratio (RS-SINR), is defined as the linear average over the power contribution (in [W]) of the resource elements carrying cell-specific reference signals divided by the linear average of the noise and interference power contribution (in [W]) over the resource elements carrying reference signals within the same frequency bandwidth. For 4G: - RS-SINR determination is based on cell-specific reference signals. For 5G: - SS-SINR determination is based on secondary synchronization signals - CSI-SINR determination is based on CSI reference signals The reference point for the RS-SINR is the antenna connector of the UE. If receiver diversity is in use by the UE, the reported value shall not be lower than the corresponding SINR of any of the individual diversity branches. If higher-layer signalling indicates certain subframes or blocks for performing RS-SINR measurements, then RS-SINR is measured in the indicated subframes. For 5G, for frequency range 2, CSI-SINR shall be measured based on the combined signal from antenna elements corresponding to a given receiver branch. For frequency range 1 and 2, if receiver diversity is in use by the UE, the reported CSI-SINR value shall not be lower than the corresponding CSI-SINR of any of the individual receiver branches.

A high-level signaling flow for activating automated data collection of key radio performance indicators, according to MDT framework, is shown in Figure 17: MDT procedural steps.

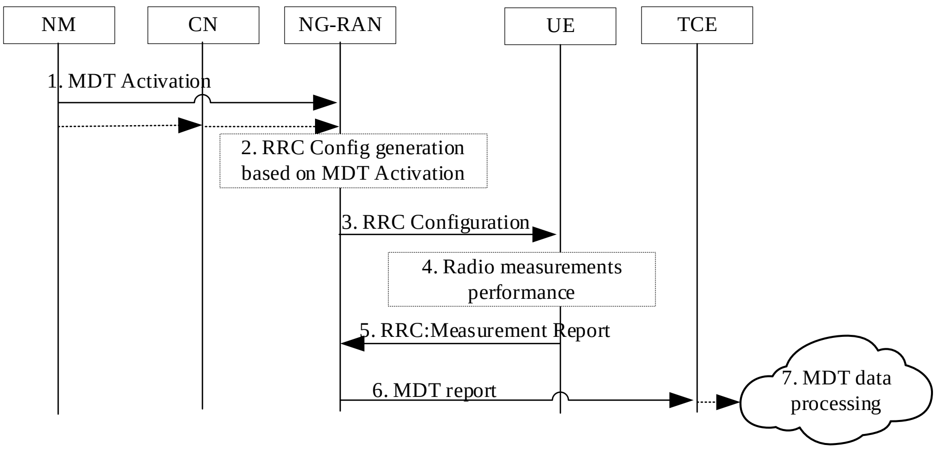


Figure 17: MDT procedural steps.

1. The Network Management (Operation and Maintenance centre of network operator) sends an MDT Activation message to the gNB. (Alternatively, the message can go through the Core Network, in case the activation targets one particular user) [30].

2. The gNB translates the configuration to the radio interface protocol and prepares the configuration, including MDT configuration parameters towards end user.

3. The gNB transmits RRC Configuration parameters consisting of RSRP, RSRQ and SINR reporting conditions (e.g. periodicity).

4. The UE (that is assumed to be active and in ongoing connection) performs the radio measurements accordingly.

5. The UE reports the measurements back to the gNB according to the predefined reporting conditions.
6. The gNB forwards the measurements to the entity outside RAN (Trace Collection Entity), for offline post-processing.
7. The data processing methods is out of scope.

The key value of the MDT automated data collection in RAN is to facilitate radio measurements collection from commercial UEs, leaving the process of data analysis to operators' choice, and 3rd parties.

3.1.2 Research developments

At the close of the 3GPP developed techniques, there has been a remarkable shift (for 4G and 5G) towards studies on feasibility of monitoring end-to-end service performance, through radio parameters.
The key radio measurements: RSRP, RSRQ, RSSI, SINR (referred to its quantized and scaled version of CQI) are considered, in [55], as four key metrics of RRM functions, which are performed for a whole cell, include packet scheduling, admission control and load control, whereas the connection based functions such as power control and handover control are executed per connection basis. The study presents thoroughly analysed A comparative observations among RSRP, RSSI and RSRQ, and relationships between SNR and RSRP based on practical measurement results recorded from a live LTE network of Australia using a commercial measurement tool NEMO Handy. The measurements were conducted while the device was moving from an area with good signal quality to the cell boundary where there is no adjacent cell to handover. A comparative analysis among RSRP, RSSI, RSRQ and SINR are provided to verify the possible dependencies among them and the effects on the ongoing connection. The study explores relationships among the measurements, such as the proportionality of RSRP and SINR, and the impact of SNR on throughput. It highlights that higher SINR correlates with increased throughput and that the difference between RSRP and RSSI affects RSRQ quality. The paper also examines network handover events, demonstrating a successful handover process based on the comparative analysis of RSRP and RSRQ values, with a focus on maintaining call quality during transitions. The findings align with theoretical principles regarding link adaptation and resource allocation in LTE networks. The findings conclude that higher SINR correlates with higher throughput during measurement slots, RSRP and SINR are generally proportional to each other on average and a smaller difference between RSSI and RSRP results in better RSRQ.

Similar set of radio metrics: RSRP, SINR, CQI, with MCS (modulation and coding scheme), and in addition physical layer throughput was subject of series of experiments in study on 'Comparing RSRP, CQI, and SINR measurements" conducted in [56]. For the purpose of the study, an indoor test network (new autonomous robotic system) was built with up to three radio points to investigate coverage and quality. The wireless measurement system consists of a standard commercial 4G mobile phone device (UE) that has been laboratory calibrated for the testing purposes, although the test was limited to only one UE (device that is using the entire bandwidth of the system) and a single base station.

For post-processing of the measured data, a post processing method worked out according to [57] and specialized software were used to allow data to be imported into the AWE tool [58]. It is worth noting the study milestone (comparing to the standardized MDT approach for offline post-processing) is the presentation of the actual results of the collected data, thus, the effect of post-processing. In context of key radio performance indicators, the relevance of the work is that it presents measurements observations comparing analysis of the collected data in coordinated and uncoordinated small cell networks. It highlights that coordinated small cells, functioning as a unified entity, significantly enhance coverage quality as measured by RSRP, SINR and MCS. The results indicate that coordinated small cells provide a more uniform signal strength and quality, with less interference in overlapping coverage areas, compared to the performance of standalone small cells which experience significant interference and poorer signal quality in the same regions. The coordinated setup resulted in higher average SINR values and improved overall user experience, particularly in areas where interference would typically degrade performance. The experiments also confirm that coordinated small cells eliminate inter-cell interference, providing high-quality, uniform coverage. Observed SINR levels were consistently above 16 dB across the measurement area, highlighting the effectiveness of the coordinated approach. The experimental validation indicates significant improvements in coverage quality for coordinated small cells based on RSRP, SINR, CQI, and MCS metrics, proving two aspects: these data types meaning in the network performance assessments and the importance of controllability.

Following the score given to RSRP, as a crucial metric for mobile network performance observability and management, studies referenced in [59] place particular emphasis on this indictor. The work investigates in detail the impact of different RSRP measurement bandwidths on the accuracy of inter-frequency handovers and overall system performance. It takes a note that RSRP measured by the users is sent to the base station and is utilized in “a hard handover algorithm”. The algorithm derives a decision whether the handover is triggered based on a negotiation with the user and target base station, to which the user would be moved for continuity of a service. The experiment has been performed using a dynamic simulator, which simulates UL and DL directions simultaneously. Various RSRP measurement configurations were tested with varying handover margins, timers, and measurement averaging implemented to minimize ping-pong handovers. The simulation methodology included various scenarios with differing velocities and measurement gap periodicities, assessing handover performance metrics like the number of handovers and spectral efficiency. The work highlights that while the 3GPP has set 4G standards for RSRP measurements using a bandwidth of 6 Physical Resource Blocks (PRBs), broader bandwidths could theoretically enhance measurement accuracy and reduce handover times. However, subsequent experiments conducted through simulations, reveal that wider bandwidths provide only marginal improvements in measurement accuracy and do not significantly advance the handover process, when no filtering over consecutive measurement gaps is applied. It is deduced that prolonged RSRP filtering periods can help to reduce the number of handovers without negatively impacting spectral efficiency. The document concludes that RSRP filtering over consecutive measurement gaps is crucial to manage handover frequency effectively without adversely impacting other performance metrics.

The role of RSRP in finding Radio Link Failures (RLF), in context of autonomous drive tests and self-optimizing solutions (elaborated in the sub-clause 1.1.4 of this dissertation i.e., SON and MDT) is particularized in [60]. The document primarily focuses on optimizing coverage in mobile networks using MDT with an extended Radio Link Failure reporting. The study highlights the importance of RLF classification in improving network performance and reducing operational costs while emphasizing the capability of this approach to inform network optimization tools effectively. The findings suggest that the implementation of this enhanced reporting mechanism can lead to significant improvements in coverage optimization and overall network performance without the need for frequent manual testing. The essential finding out of this study is that: RSRP is a key measurement that can help in determining the cause of RLFs, specifically whether they are due to coverage issues, handover parameterization, or interference problems. Therefore, the metric is also vital for coverage optimization (low RSRP values indicate weak coverage areas). Its provision allows network operators to target specific areas for coverage improvement, thereby reducing RLF occurrences. Yet, the study proves that adjusting handover parameters based on RSRP data can enhance network performance, and thus justifies RSRP use to network decisions for handover, which aligns with findings in [59]. The paper analyzes the effectiveness of RSRP reporting, through dynamic system-level simulations, demonstrating that if it is included as component of RLF report it can accurately pinpoint coverage problems with minimal standardization efforts. Conclusively, in context of the automated solution, it is explored that the RSRP-related data from extended RLF reporting feed into the broader MDT and (SON) frameworks, can allow for more accurate and automated network optimizations over time.

Another example, focusing on RSRP as a key quality metric observation in 5G specific scenario is outlined in [61]. This research utilized a novel Matlab-based simulator to analyze handover behavior in dynamic environments with a target to estimate and limit undesired handovers in mobile networks. Pre-requisite experiment was carried out in real network in the industrial area of Reggio Emilia (RE), Italy for actual data collection. The used experimental testbed composed by a Sierra Wireless EM9191 5G embedded wireless modem hosting a Subscriber Identification Module (SIM) card of the 5G provider of the considered gNB, and connected to an antenna using a SubMiniature version A (SMA) cable. With this setup, CSI-RSRP measurements were collected in real environment and derived to a Matlab-based environment relying on its 5G Toolbox to increase the simulator accuracy. In the performance analysis, the network performance is investigated in terms of CSI-RSRP used to express the gNB-UE link channel quality with the split to two types of CSI-RS defined: Non-Zero Power (NZP) CSI-RS and Zero Power (ZP) CSI-RS. In detailed analysis, NZP CSI-RS is used at the UE for received power measurements for beam and mobility management, DL CSI acquisition, interference measurement, and time and frequency tracking. The work highlights the challenges posed by multiple, often unnecessary handovers due to the high density of gNBs in urban areas, which can degrade QoE and lead to increased latency and reduced throughput. It details RSRP changes over time and as a function of the distance between the UE and the gNB. It concludes with a focus on the importance of effective handover management algorithms to improve connectivity and reduce the frequency of unnecessary handovers, suggesting future research directions to validate the simulator in varied conditions and develop predictive mechanisms to enhance QoS and QoE for mobile users.

A comprehensive analysis of the coverage and performance of joined 4G and 5G deployments is elaborated in [62]. The study is motivated, by claimed presently, 5G networks deployment in two modes: Standalone (SA), where operators use only 5G for the entire network stack, and Non-Standalone (NSA), which uses a 4G channel as a primary anchor channel thus easing the transition from 4G to 5G SA, and secondly: largely under-explored observability of 5G commercial deployment impact due to mmWaves. The document presents throughout overview of deployments and propose scalable methodology for understanding real-world network performance and addressing ongoing challenges in 5G technology. The employed in the study experiment is focusing on various frequency bands (low, mid, and high) and their respective trade-offs regarding coverage, throughput, and latency. 4G and 5G signal analysis is based on data records from SigCap application enabling collection of RSRP and RSRQ for both: 4G and 5G. Although, there are challenges noted of measuring real-world 5G deployments with quantifying 5G mmWave performance, the experiment reveals notable decrease of RSRP by 10 dB once the device is under vulnerability of mmWave connections, while RSRQ remains insensitive (at the same level). The explored methods and studies on the measurements conclude the fundamental finding that even with co-existing with 4G deployments, 5G connections exhibit considerable potential in its early stages, and additional optimization efforts are required to guarantee that the requirements of 5G are met. Alternate conclusion drawn from the work can be that RSRP provides more reliable insights, compared to RSRQ.

Alternative study, placing SINR in focus is detailed in [63]. The document presents a study on adaptive beamforming techniques aimed at maximizing the signal-to-interference-plus-noise ratio (SINR) in scenarios with correlated interferences. It critiques conventional methods like the minimum variance distortion-less response (MVDR) that suffer from performance degradation due to signal cancellation when facing correlated interferences. The study proposes a new beamforming method named the Composite Steering Method (CSM), which utilizes a composite steering vector (CSV) and an interference covariance matrix (ICM) derived from uncorrelated components to achieve optimal SINR. The methodology involves estimating the CSV and ICM from sample matrices using oblique projections while avoiding the direct estimation of interference and desired signal powers. The document uses SINR as a reference to adjust network performance, concluding there is future research required on optimal adjustments and achieving better SINR in applications needed for adaptive beamforming.

As the eMBB, mMTC, and URLLC are the three primary capabilities designed for 5G networks as classified by ITU, its worth noting the research developments that address the unique attributes and needs of eMBB, mMTC, and URLLC. Exemplary study in [64] is based on some simplifying assumptions, however, it is stated the basic model and the methodology developed here can be extended to more general models and other operations under joined 5G scenarios. Nevertheless, the study introduces Heterogeneity Non-Orthogonal Multi-Access (H-NOMA) for resource sharing among devices from different slices and with diverse requirements. H-NOMA contrasts with traditional Non-Orthogonal Multiple Access (NOMA) by allowing multiplexing of different services rather than homogeneous ones. Dependability heterogeneity is proposed as a model paradigm that utilizes varying reliability needs across applications. Starting with eMBB-URLLC coexistence, H-NOMA scheme is applied in a way that eMBB and URLLC users are allowed to access partially non-orthogonal resources, so that only a subset of frequency channels potentially occupied by URLLC traffic may be interfered by eMBB transmissions. Another direct generalization undertaken is to assume that the minislots are pre-allocated to different URLLC devices. If each URLLC transmission is carried out by a different device, then then error decoding events are independent across the mini slots. As a more involved extensions of the model, it is proposed to consider the impact of frequency diversity also for eMBB traffic, and the performance under alternative decoding strategies, such as treating interference as noise. What the research develops is a communication theoretic model that accounts for the unique attributes and needs of eMBB, mMTC, and URLLC. Efficiency trade-offs between the three services are examined, showing potential gains from non-orthogonal slicing in specific conditions. The study highlights that assumed model for slicing often provides substantial performance advantages in scenarios with high eMBB rates and low interference from mMTC devices. The findings suggest that non-orthogonal slicing can improve performance when dependability variety is effectively utilized. Ultimately, the findings conclude that H-NOMA can outperform traditional methods under certain conditions by effectively managing the diverse demands of the three services types and non-orthogonal resource assignment can lead to improved efficiency trade-offs compared to traditional orthogonal slicing methods, which according to design supports a high degree of flexibility for configuration [65].

The co-existence of 5G major service classes: URLLC and eMBB is presented in study in [66]. The authors propose a null-space-based spatial preemptive scheduler (NSBPS) designed to ensure that URLLC traffic, which requires stringent latency and reliability, can coexist with eMBB traffic, which demands high data rates. The NSBPS scheduler aims to minimize queuing delays for URLLC applications while maximizing the ergodic capacity for eMBB users by utilizing a spatial subspace for interference-free transmission of URLLC packets. The paper includes analytical analysis and extensive system-level simulations to validate the performance of the proposed scheduler against existing solutions. Results indicate that the proposed scheduler provides robust latency performance for URLLC while significantly improving eMBB capacity compared to state-of-the-art scheduling methods. The findings highlight the critical balance needed in resource allocation among the competing requirements of both service classes in densely populated 5G networks. In addition, the study emphasizes the importance of user-centric scheduling approaches in achieving optimal performance in future mobile networks.

3.2 ML Applicability to Mobile Networks

3.2.1 3GPP standard developments

3GPP as international standardization body is responsible for setting the overall policy and strategy of cellular telecommunication network development, by approving and maintaining scope of the technological advancements. The technologies include radio access, core network, and service capabilities, which provide a complete system for mobile communications. Although, 5G has been hailed to bring a major paradigm shift with life digitalization, the initial scope for 5G specifications (3GPP Release 15) didn't consider AI applicability [67]. The AI revolutionizing all major industry sectors, impacted 5G roadmap, broadening the 5G-Advanced of 3GPP technology to applications including ML enablers. These integration of ML into 5G systems is currently under exploration and covers several threads:

i. AI/ML-related requirements for the 5G system – defining generic and high level requirements for 5G system to facilitate ML integration [68], including detailed study on requirements associated with AI/ML model transfer [69]
ii. AI/ML Management for 5G Systems – defining use cases for 5G and future generation mobile networks that will require AI/ML capabilities, with detailed split into:
   - Data Analytics Management (MDA)– with an objective to work out a generic mechanism for managing the ML training for any kind of AI/ML enabled capabilities [70], [71].
   - Generic management of the AI/ML capabilities in 5G system, addressing more than 40 use cases related to ML process, such as AI/ML orchestrating, ML entity control and monitoring coordination between ML capabilities [72], [73].
   - 5G system support for AI/ML-based services- with the scope to identify and validate which 5G system domains are relevant for applying ML, such as service provision, and resource optimization and what high level policies should be applied [74].
iii. AI/ML for Radio Access Network, covering exploration on how to integrate ML into RAN operations with detailed insights and target scattered per target use cases:
   - Network Slicing and Coverage and Capacity Optimization [75], which are subsequent scenarios after having the same flavour of operations established for Network Energy Saving, Load Balancing and Mobility Optimization [50]. All being the SON optimization targets with RAN intelligence enhancements through ML enablers.

iv. AI/ML for NR Air Interface, targeting data collection framework and necessary procedures in radio interface to support ML support for two radio use cases: Beam Management Optimization and Positioning [76].

By developing the AI impact to 5G system, network management, resource allocation, service quality or type, the standardization efforts are focused on finding requirements that will facilitate the ML integration (i, ii). The efforts associated to RAN and Air interface (iii, iv) lead to definition of procedural details that need to be implemented in RAN, and therefore facilitate developed solutions feasibility. The latter scope (iii, iv) offers significant guideline in terms of integration and inter-operability between the device and base station, but does not impose specific ML algorithm use. For instance, it is guided that ML algorithms may predict the energy efficiency and load state of the next period, which can be used to make better decisions on cell activation/deactivation for energy saving. Based on the predicted load, the system may dynamically configure the energy-saving strategy (e.g., the switch-off timing and granularity, offloading actions) to keep a balance between system performance and energy efficiency and to reduce the energy consumption [50].

The activities exploring integration of AI/ML to RAN worked out an architecture that organizes an intelligence in RAN with ML specific operations as depicted in Figure 18.

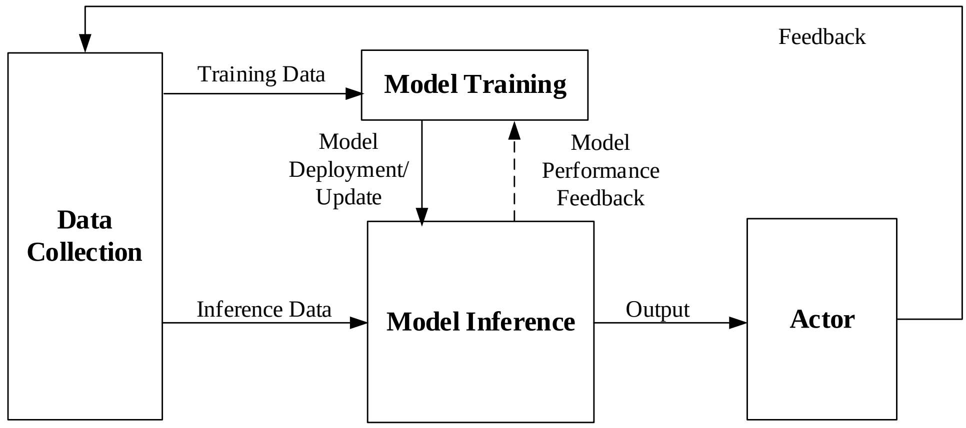


Figure 18: Functional framework for RAN intelligence, source: [50].

This functional framework for RAN implementing ML, differentiates several functions [50]:

- Data collection that provides input data to Model training and Model inference functions. The Data collection plays the role of ML algorithm specific data preparation.
- Model Training that performs the AI/ML model training, validation, and testing which may generate model performance metrics as part of the model testing procedure. The Model Training function is also responsible for data preparation.
- Model Inference that provides AI/ML model inference output (e.g., predictions or decisions).
- Actor that receives the output from the Model Inference function and triggers or performs corresponding actions. The Actor may trigger actions directed to other network entities or to itself.

The Data Collection can be located in RAN only, though can derive data from the UE inputs, and differentiate various datasets depending on the use case (Mobility Robustness, Energy Savings, Coverage and Capacity Optimization or Network Slicing) [50].

A separate consideration on Data Collection, which can be realized with a greater freedom, without imposing details on how RAN architecture integrates ML in, is conducted under work for AI/ML for NR Air Interface, addressing beam management optimization and positioning. In line with the study in [77], there are various Network-UE collaboration levels between the UE and the gNB possible, diverse kinds of recommendations for information that will enable ML-enabled procedures (e.g. best beams identification, UE location), varied further due to dependance on ML-function sidedness (e.g.: UE-side ML functionality or RAN-side ML functionality). Three levels of Network-UE collaboration were identified for AI/ML for NR Air interface [77]:

- Level x: No collaboration

- Level y: Signaling-based collaboration without model transfer

- Level z: Signaling-based collaboration with model transfer

In Level x, there is no collaboration between the UE and the network, meaning that ML operations rely entirely on implementation, with no specific effects due to this lack of collaboration. In Level y, transparency is maintained between the AI/ML model used by the UE and the network. Although the model itself is not directly transferred within the standardized 3GPP domain, supporting information for the AI/ML model can still be communicated through 3GPP signaling protocols, such as RRC messages [76], [78].

The standardization work is progressing, and related frameworks definitions continue. However, for any of the RAN-focused developments, ML methodology type (e.g. supervised, or non-supervised) and ML algorithm selection are open for operator's choice and remains beyond standardization target [44]. The aim by the 3GPP developments is mainly identification of the framework, procedures and identification of ML algorithm inputs that are considered useful for a given use case scenario.

3.2.2 Research developments

The integration of ML into 5G system has become a central focus of research activities. The literature on this topic presents a wide range of approaches, from sophisticated methods aimed at advancing radio physical layer (PHY) components and physical channel coding [79] to mature discussions on ML innovations inspired by its applicability to 5G, such as construction of Graph Neural Networks (GNN) for wireless communication [80]. The ultimate goal is to demonstrate the value of AI in modern networks, offering numerous examples of ML algorithms and revolutionary ideas that could significantly enhance network intelligence, understood generally as smart adoptability based on learning capabilities and network autonomy.

Starting from the literature in the machine learning and data science domain, its self-evident that many ML algorithms have been developed even before judging their usefulness to any of todays' emerging fields for AI applicability. To exemplify, the work in [81] overviews several Supervised and Unsupervised algorithms (e.g. Decision Trees, Rule-Based Classifiers, Naïve Bayesian Classification, k-Nearest Neighbors Classifiers, Linear Discriminant Analysis, Support Vector Machine, Principal Component Analysis), clarifying their well-known mathematical roots with comprehensive explanation how they can apply for AI. However, this is without the recognition it can become a paradigm for 5G networks. The document teaches about optimization theories, that are central to ML, where the goal is to minimize or maximize an objective function (e.g., error rates, likelihoods) by adjusting model parameters or Probability and Statistics, which provide the foundation for understanding and modeling uncertainty, which is inherent in most real-world data. Statistical methods are found crucial for model evaluation and hypothesis testing. Statistical applications, such as Bayesian networks, Markov models, and statistical inference techniques are essential in areas such as predictive modeling, data mining, and recommendation systems. Furthermore, linear algebra is referenced, as a fundamental tool to the representation and manipulation of data in ML. It underpins many algorithms, particularly in high-dimensional data processing. Principal Component Analysis (PCA) is there as a technique that transforms a set of correlated variables into a set of uncorrelated variables known as principal components. Vast amount of such broadly developed mathematical methodologies exist, and they can be understood as uniform techniques, that can become applicable to many fields, including telecommunication.

The emergence of AI and 5G creates now a shift in applying the well-known mathematical methods through ML to mobile networks optimization. As an example, [82] discusses the development and implementation of statistical algorithms for fault detection and prediction in wireless networks, focusing on the Operational Fault Detection (OFD) algorithms. It highlights the need for high reliability and availability at low costs, advocating for a proactive approach to fault management rather than a reactive one. The paper contrasts first-generation OFD algorithms, which require significant human intervention to set thresholds, with second-generation OFD algorithms that are adaptive and minimize human input. The second-generation algorithms enhance fault detection by identifying performance degradations, which are precursors to catastrophic failures. The paper details the methodologies behind these algorithms, including learning and correlation techniques to monitor system performance metrics effectively. It emphasizes the importance of reducing operational expenses and capital expenditures through automated fault prediction and early detection, thereby minimizing downtime and revenue loss. Additionally, simulation results demonstrate the robustness and effectiveness of these algorithms in real-world scenarios. The study concludes that adopting such predictive fault management strategies is essential for maintaining high availability in wireless networks at an affordable cost.

An advanced, but still based on probability and statistics theories study in [83] elaborates on existing 5G channel models for the high frequencies and with context of emerging 5G complexities, necessitating new radio propagation models. The new models must accommodate large antenna arrays, multi-band operations, and varying mobility scenarios. For that reason the research is conducted through ML-based channel modeling, using algorithms to learn from data and make predictions about future channel behavior; regression algorithms that predict variables by simulating Gaussian distribution functions for channel prediction, demonstrating high accuracy in data reception, and NN due to their capability to handle complex, non-linear relationships. The authors highlight that the new channel models must accommodate various scenarios with high mobility. By leveraging ML, particularly regression and classification algorithms, the research aims to enhance the prediction of channel behavior using data-driven approaches. The regression model simulating a Gaussian channel is promoted, due to promising results in accurately predicting transmitted data amidst noise. The paper concludes by indicating future work will involve testing various datasets and ML algorithms to refine channel modeling accuracy and computational efficiency.

Another work motivated by 5G channel perception deficiencies is in [84]. It studies a predictive beamforming using the deep learning approach in the multiuser downlink. A general framework that predicts the beamforming solution to maximize the sum rate with historical channel measurement data is proposed, with simulation results showing that the proposed deep learning-based solution achieves significantly higher effective sum rate over the traditional channel estimation-based optimization and the separate prediction and then optimization scheme. To the existing modelling based on channel estimation error that follows the complex Gaussian distribution with zero mean and variance matrix, the method is using novel design with a loss function and introducing the attention mechanism to allow the algorithm to put focus on different historical channel.

Worth noting that completely opposite view, promoting methods that can enable ML without the need to rely on accurate mathematical models, can be found in [85]. It claims Deep Learning is an emerging interdisciplinary paradigm, by a comprehensive review of the application of Deep Learning (DL) in wireless communications, highlighting its potential to address emerging challenges brought by new applications. It discusses two primary methodologies: DL-based architecture design and DL-based algorithm design. The former involves reformulating traditional block-based communication design using DL, exemplified by advancements in receiver and joint transceiver design. The latter focuses on enhancing algorithmic efficiency and performance through DL, with examples including channel estimation and decoding techniques for 5G systems. The review emphasizes the advantages of DL, such as its ability to learn complex relationships, making it suitable for complex communication scenarios. It promotes DL-based algorithms as in contrast to classical methods, under the umbrella of Shannon’s information theory, DL-based methods do not require solid mathematical foundations in terms of theoretical analysis. Given, the optimization plays an important role in wireless communication systems to realize efficient exploitation of limited radio resources, many ML optimization algorithms introduce counter-productive solutions. They require a large number of iterations to converge, which results in both high complexity and high latency, especially when the problem’s scale is very large. DL-methodology is promoted to tackle these challenges. The study shows that DL can be used to speed up processing while maintaining reliable performance.

In context of the AI applicability to RAN, the work in [86] provides a remarkably comprehensive survey of 15 years of literature on ML techniques applied to SON. The study notes that numerous 5G requirements and resulting breakthroughs are very important and often referred as a necessity for future mobile networks, but will require heavier changes in the network and possibly a change in paradigm in terms of how network solutions are provided, therefore enhanced SON intelligence with a remedy internally in the network is advantageous choice. The surveyed methods and algorithms applied to SON solutions vary from basic control loops and threshold comparisons to more complex mathematical, ML and data mining techniques. The variety of ML techniques is classified to address specific functions such as self-configuration, self-optimization, and self-healing. Some of the key approaches differentiate:

- Supervised Learning: which is training a model using labeled data, where the algorithm learns to predict an output based on the input data. This method is particularly useful in scenarios where the expected output is known, allowing the system to learn the mapping from input to output. The most common algorithms recognized under Supervised Learning are:

i. Bayes' Theory: working on basis of an important rule in probability and statistical analysis to compute conditional probabilities.
ii. k -Nearest Neighbor (k -NN): with an algorithm applicable to problems where the underlying joint distribution of the observation and the result is not known.
iii. Neural Networks (NNs): with NN approach, that is based on the assumption the connections between different layers always go forward and do not form a cycle, therefore, this type of network is commonly known as feed forward neural network. There are other types of NNs, but this paper focuses only on feed forward NNs
iv. Support Vector Machine (SVM): with an idea to map a set of inputs into a higher dimensional feature space. This is done through some linear or non-linear mapping and its objective is to maximize the distance between different classes. Since the goal of SVM is to find the hyperplane that produces the largest margin between different classes.
v. Decision Trees (DT), which are constructed by repeated splits of subsets of the original data into descendant subsets.
vi. Recommender Systems: also known as Collaborative Filtering (CF).

- Unsupervised Learning: This method is used when the output is not known, and the system must find patterns and structure from the input data on its own. It is often used for clustering and anomaly detection in network operations. The most common algorithms recognized under Unsupervised Learning are:

i. K -Means: One of the most popular unsupervised learning algorithms. This clustering algorithm is very useful in finding clusters and its centers in a set of unlabeled data. The algorithm is very simple and only requires two parameters: the initial data set and the desired number of clusters.
ii. Self-Organizing Maps (SOMs): Another popular clustering method, which is a that attempts to visualize similarity relations in a set of data items.
iii. Anomaly Detectors: involving Anomaly Detection (AD) techniques, with a main goal to identify data points that do not conform to a certain pattern observed in the data.

- Controllers: with a note that controllers do not belong to the class of intelligent algorithms, the study teaches they have been extensively used to perform basic SON tasks in cellular networks due to their simplicity and ease of implementation. They split to:

i. Close Loop Controller: Also known as Feedback Controllers, rely on a feedback mechanism between the input and output in order to constantly adjust its parameters.
ii. Fuzzy Logic Controllers: In contrast to normal feedback controllers, that use classical logic (Boolean logic), these controllers use fuzzy logic, a type of logic that represents partial truths. This process is done by applying an interpolation between the two extremes of binary logic (0 and 1). Since these controllers have a better granularity than standard binary logic controllers, generally, more detailed and complex solutions can be achieved by FLCs than feedback controllers.

- Reinforcement Learning (RL): RL involves training models through a reward system, where the algorithm learns by interacting with the environment. Correct actions are rewarded, while incorrect ones are penalized, helping the system improve over time. This method is highly applicable to dynamic network environments where decisions need to be continuously optimized.
- Heuristic Methods: These are problem-solving approaches that use a practical method not guaranteed to be perfect but good enough to reach an immediate goal. Heuristic methods in SON include brute-force search and metaheuristics, which are higher-level strategies designed to generate solutions that are sufficiently close to the best possible solution.
- Genetic Algorithms (GA): Inspired by the process of natural selection, GAs evolve solutions to problems through mechanisms akin to biological evolution, such as selection, mutation, and

crossover. GAs are applied in SON for tasks like radio parameter configuration, coverage and capacity optimization, and load balancing.
- Dimension Reduction Techniques: These are used to reduce the complexity of data and improve the performance of algorithms. Techniques like Principal Component Analysis (PCA) and Multi-Dimensional Scaling (MDS) are used in SON to simplify the input data while preserving its essential structure.
- Markov Models: These techniques are proposed for SON to manage uncertainty and model the probabilistic nature of network states and transitions, respectively.

The study also proposes few important aspects to consider ML algorithms applicability:
- Scalability: as a concept determining if an algorithm is able to handle an increase in its scale, such as feeding more data to the system, adding more features to the input data or adding more layers in a NN, without it limitlessly increasing its complexity. In order to cope with future networks, which are expected to be much more dense and generate much more data, scalability is a highly desirable feature so that algorithms can be deployed easily and quickly in the network. The scalability can also help in determining if certain types of algorithms can be mass deployed in decentralized or centralized networks.
- Training time: as a metric that represents the amount of time that each algorithm takes to be fully trained and for it to be able to make its predictions.
- Response time: related to the agility of a system is the response time of an algorithm.
- Training data: the amount and type of training data an algorithm needs.
- Complexity: which can be defined as the amount of mathematical operations that it performs in order to achieve a desired solution. Complexity also relates to the power consumption of a system, as a system that needs to perform more operations will, consequently, need more power to operate. Hence, this concept can determine if certain algorithms are more suitable to be deployed at the user or operator’s side.
- Accuracy: as a parameter determining caching the right content, with quick mitigation of the impacts to the network.
- Convergence Time: which relates to the time an algorithm takes to make a prediction.
- Convergence Reliability: parameter of learning algorithms in the initial conditions that they are set in and their convergence reliability.

Overall, the survey is not conclusive, and notes there are still lots of ML algorithms that have not been applied to certain SON functions, suggesting further exploration of ML solutions still need to be done in order to investigate their performance and determine if these methods can really work or not.

Motivated strictly by 5G specific scenario: URLLC, increasing demands of IoT applications, the research in [87] promotes focusing on ANN implementation in wireless network environments. The study claims, that the development of ANN-based solutions for IoT ecosystem will be crucial for the 6G wireless systems. The document is a comprehensive tutorial on different types of ANNs, including recurrent, spiking, and deep neural networks, along with their architectures and relevant applications in wireless networking. It highlights specific use cases such as drones, virtual reality, IoT devices, mobile edge caching and computing, and multi-radio access, SON technologies, VR, HD streaming and for each application, it provides motivation for using ANNs, outlines the associated challenges, and presents detailed examples of practical implementations. By the numerous examples and applicability, it emphasizes the role of ANNs in enhancing data analytics capabilities, enabling real-time resource management, and improving user experience through predictive analytics. Finally, the document emphasizes that by underscoring the significance of integrating machine learning and ANN techniques into wireless network design, the full potential of future networks is still to explore and pave the way to improved communication efficiency.

Yet, IoT traffic with billions of connected devices claimed, as presenting significant challenges in resource management (RRM), became a key research area in [88]. Due to massive channel access, requirement to cope with sophisticated techniques to manage interference and power allocation. The study finds there are much more challenges resulting from this 5G specific usage scenario: i. Harmonious Coexistence of IoT and Human-to-Human Traffic, ensuring that IoT traffic doesn't disrupt human communication networks; ii. the need to provide coverage extension, as IoT devices with limited power need more detailed planning for Device-to-Device (D2D) communication, and extended coverage; iii. energy efficiency management for the limited battery life of IoT devices; iv. real-time processing and low latency, for mission-critical IoT applications; v. heterogeneous QoS to cope with diverse devices with varying capabilities and requirements. The limitations of traditional resource management techniques, often fall short in dynamic IoT environments due to their complexity, lack of real-time capability, and inability to handle massive data. The work motivates ML, as well suited for scenarios where traditional methods struggle, particularly when data-driven, real-time, adaptive solutions are required, since ML can handle the complex, dynamic nature of IoT networks, offering lower complexity and better contextual decision-making. To justify the statement, the study conducts a survey through the existing literature. The survey focuses on ML and Deep Learning techniques use in complex network environments like D2D, HetNets, MIMO, and NOMA. It is concluded that there is notable record to claim ML allows for dynamic adaptation to changing network conditions, making it suitable for environments with diverse user requirements: ML techniques can help optimize resource allocation by leveraging vast amounts of data generated by IoT devices, enhancing automated decision-making and Deep Reinforcement Learning (DRL) is particularly effective for resource allocation problems, providing solutions to complex, high-dimensional challenges.

IoT is known as one 5G usage scenario or traffic type. In contrast to very use case specific consideration, more generic framework for dynamic slice management, allowing efficient planning and scheduling of network resources is proposed in [89]. The Intelligent-Slicing framework aims to optimize network slicing for 5G-and-beyond networks, addressing diverse communication and computation requirements. It leverages Network Function Virtualization (NFV) to collect service requirements and map them to network management decisions and supports various KPIs including reliability, energy consumption, and data quality. BY this differs from traditional metrics, focusing on high-quality performance rather than just throughput and latency. The method proposes AI Integration to the 5G network through an AI-assisted Network Function Virtualization (NFV) and Software-Defined Networking (SDN). It introduces references for planning and scheduling phases of network slicing, utilizing reinforcement learning (RL) and distributed optimization techniques. The results demonstrate the effectiveness of the proposed framework in improving reliability, energy consumption, and data quality compared to existing methods. The integration of AI is claimed to be necessary in managing network slicing throughout its entire (AI) lifecycle. The framework promotes also utilizing AI across different network tiers, from user devices to the cloud.

Another, hybrid approach for evaluating QoE for potentially all 5G services within 5G SA networks, which considers jointly analytical and ML methodologies can be found in [90]. It combines flow-level packet inspection with machine learning to enhance service detection and QoE measurement. The approach aims to reduce operational costs associated with legacy monitoring techniques in high-traffic 5G environments. Based on the motivation that traditional QoE and QoS monitoring systems are inadequate for the complex architecture of 5G networks, and remain costly and ineffective due to the volume of user-generated traffic in 5G environments, the study proposes a framework that uses deep packet inspection (DPI) to capture and analyze raw user traffic for service classification and QoE assessment. On top, ML algorithms are employed to detect and categorize eMBB services from previously unresolved traffic. The study was conducted on a live 5G SA network, demonstrating successful real-time traffic analysis and service identification. Among six ML models tested on the flow dataset, the best model is selected based on the performance evaluation metrics (i.e., XGBoost algorithm). The work concludes that cost-effective and scalable monitoring solutions for 5G networks remain a critical area for further exploration.

3.3 Literature overview conclusion

Based on the literature review, it can be summarized that:

- Most studied, essential and deemed network optimization methods consist of radio access network (RAN) optimization process. This process can be facilitated through the utilization of end user measurements such as RSRP, RSRQ and SINR.
- Conducting drive tests is a continued trend in providing understanding on network performance optimization methods, justifying existing in radio optimization solutions or motivating any further advancements. It remains limited and unexplored on the methodology for offline post-processing of typical drive-test entries, leaving the process unresolved by standardized requirements.
    - Compared to 4G, practical observations on 5G network performance in real-world, including both: use case and traffic specific scenario or generalized observability for any traffic, slice or deployment type are under-explored and new paradigm due to ongoing 5G networks rollouts.
    - Co-existence of eMBB, URLLC, mMTC introduce challenges to tackle the user-centric requirements by appropriate balancing of operations on the network side.
- The research trend for ML applicability to wireless networks is under evolution in industry and academia. While standard-driven directions do not consider ML algorithms explanation currently, leaving the choice to operators, research-driven literature records numerous methods approaching deep learning and machine learning technics utilization for 5G networks optimizations. The techniques are scattered and various depending on a use case severity. The breakthroughs in scientific literature are scaling from statistical and simple methods, to advances requiring sophisticated and complex mathematical modelling embedded into neural networks paradigms, that creates an implementation challenge. There is no conclusive nor converging direction found that could guide network implementations to achieve comparable results.
- Future research directions conclusively indicate a potential and need for applying known ML benefits (e.g. predictiveness and self-learning) into network optimization methods.

3.4 Identified cognitive aps in the literature

The existing prior art records ever-evolving methods for radio network observability and its performance improvements with an ultimate target to achieve end-user satisfaction. The network optimization methods continuously evolve, under considerations of critical data collection that play a crucial role in the radio domain and user satisfaction assessment. The origin for dataset collection is relatively constant and determined: to use key performance indicators from the user as primary units. The most advanced technologies considered for processing the data on the network side, apply AI. The literature records numerous methods for integrating ML algorithms and statistical methodologies in RAN operation, based on the assumption that data collection is provisioned per specific use case.

There is however lack of AI applicability to MDT-powered data collection, which enables retrieval the most basic and key radio performance metrics from any 5G device. Therefore, the remaining gap is in applying ML algorithm to autonomous data collection facilitated by MDT and empowering the MDT with self-learning and predictiveness capabilities.

The dissertation addresses this by experiment in the next chapter.

4 Methodology

4.1 Methodology Design

To address the research problem on how ML can be effectively applied to RAN optimization, focusing on radio monitoring without a need to introduce additional overhead, the proposed strategy is to model ML-based technique for 5G RAN performance evaluation that interacts with the radio interface environment. It should allow an Actor (i.e., RAN) to observe and learn radio interface measures. The observations based on the input data received from the end user, which are measurements provided by the users' devices operating in live-network, should be accessible through the interactive information exchange. To enable the ML paradigm, it is proposed to construct ML model with data processing actions. The ML-model use decision tree-based evaluation, with ANN algorithm adoption for parallel processing of few dimensions (e.g. different performance indicators) The actions should derive ML model outcome, which can lead to decision making function on scaling data collection (e.g.: lead to reconfiguration or measurements reporting filtering).
The proposed ML scheme envisions the following steps (also depicted in Figure 19):

i. Input Data provision to the Actor (i.e., RAN node), where:
    - Data is derived in a parameterized form
    - The parametrization classifies multivariate data sets to dimensions related to exploiting 5G traffic and RAN performance.

ii. Processing Input Data to derive input to ML algorithm in a form of labelled data.

iii. Modelling ML operation to perform data observations and learning actions.
    - The incorporated ML algorithm consider simple statistical methods (such as linear regression) or ANN-based algorithms adoption for parallel processing of different Input Data

iv. ML model operation evaluation of the outcome based on the multiple criteria, where:

v. The evaluation aims at trade-off between complexity and measurable gains

vi. Decision making function triggering the Actor to perform the Input Data parametrization.

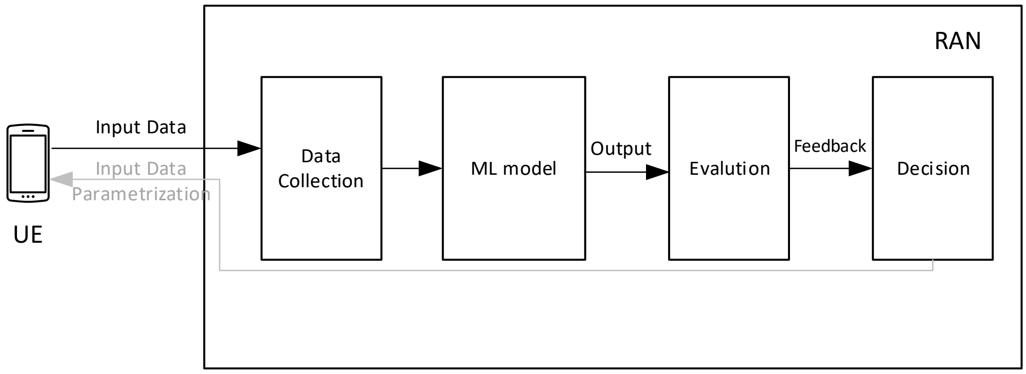


Figure 19. High level scheme for applying AI/ML-based 5G RAN performance evaluation based on 3GPP-assumed paradigm.

To exploit the methodology feasibility, this dissertation undertakes the following steps (as illustrated in Figure 20):

I. Data Collection through Drive tests focusing on eMBB use case and 5G NSA deployments, with the following test scenarios covered:
    a. two MNOs in urban area (Wroclaw, Poland)
    b. one MNO in rural area (Mazovian, Poland)
    c. one MNO in highway (Lower Silesian Voivodeship)
    d. one MNO in urban area (The Island Continent)

II. Data analysis targeting findings on radio metrics trends and their relationships, with the aim to identify essentiality of the Input Data for Data Collection qualified for ML model
III. ML method application to the qualified data
   i. With three supervised ML algorithms applied and tested with diversified datasets, where radio specific indicators from access networks are used as input, and findings from data analysis (step II) are used as a guiding policy
IV. Finally, the work analyses ML methodology results towards aiding the utilization of ML in RAN optimization.

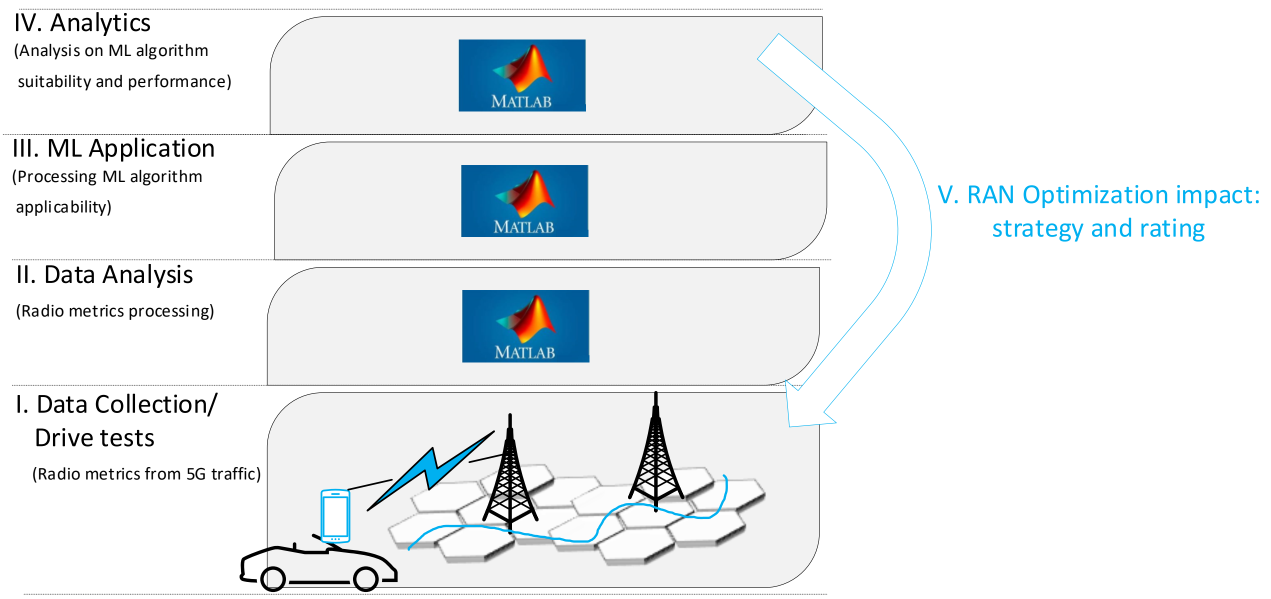


Figure 20: High level scheme for methodology undertaken in the research.

4.2 Experimental Data

4.2.1 Data Collection

There are many drivers behind the anticipated importance of the 5G characteristics, dependant on the detailed circumstances. To select the 5G traffic specific indicators for ML-driven Data Collection under this work, the strategy is to first understand the practical scenarios in the the current 5G rollout phases, availability of the 5G services and implications of the 5G network setup. To grasp the 5G network accessibility and practical effects of the 5G framework, the empirical research aimed mainly at uncovering how deployment and technological characteristics change based on a perception by the 5G-capable commercial devices (Nokia 8.3 5G, NokiaG42 5G and Samsung SM-S906E, as presented in Table 8). Gathering data sets has been carried out in outdoor areas through manual drive test in four different geographical regions, under:

- **5G NSA deployments**:
  a. for two MNOs and two devices in urban area (Wroclaw, Poland)
     - MNO1 bands:20EUDD 803.5, 3DCS 1832.4, 7IMTE 2680, 1IMT 2162.5 [3]
     - MNO2 bands: 20EUDD 796, 3DCS 1859.9, 1IMT 2117.5

[3] According to bands' notions displayed in the application used for data collection

b. for one MNO1 and two devices in sub-urban and rural area (Mazovian Voivodeship, Poland)
   - Device1: 20EUDD 803.5, 7IMTE 2680.0, 3DCS 1832.4, (1IMT 803.5 shortly)
   - Device2: 20EUDD 803.5, 3DCS 1832.4
c. for one MNO1 and one device in highway (Lower Silesian Voivodeship)
   - MNO1: 3DCS 1832.4, 1IMT803.5, (20EUDD 803.5 shortly)
d. two MNOs and one device in urban area (The Island Continent)
   - MNO3: n28 700, n78: 3300-3800 (TDD)
   - MNO4: ‘E-UTRA’ (as recorded in log records)

- **Campaigns schedules**:
  - Under regular mid-day load, without huge congestions or traffic density, whereas covering routes along more busy and less busies areas
  - Under different mobility scenarios: for urban and rural area with varying driving speed 0-40km/h, while for highway area with average speed: 120-140km/h
- **Measurements setup:**
  - key radio performance indicators were recorded with ‘Network Cell Info’ Android application in test case scenarios a, b and c. Each data record from the application consists of device identifiers (sim id, device type), serving MNO (operator name), network setup (radio type, carrier, area id, cell id, mobile country code, detailed location information (geo-coordinates) of the device, time stamp, RSRP, RSRQ, SINR, accuracy, speed)
  - the application was recording samples every 2 seconds, and was running simultaneously on each of the devices in test a and b, or on a single device in test c
  - specifically for MNOs abroad (test d), different toolkit has been used for logging data, with capability to record additional attributes: NR PCell Throughput and Time to First 500Kbytes, in addition to RSRP, RSRQ and SINR
  - samples sets were varying from 700 to 14000 samples in tests a-d
  - signal transmissions were dependant on network frequency bands (as shown in Table 9).

Table 8: Devices characteristics.

	5G bands	Chipset version
Nokia 8.3	n1, n2, n3, n5, n7, n8, n28, n38 (2600MHz), n40, n41, n66, n71, n78 (3500MHz)	Qualcomm SM7250 Snapdragon 765G 5G (7 nm)
Nokia G42	n1, n2, n3, n5, n7, n8, n28, n40, n41, n66, n77, n78	Qualcomm SM4350-AC Snapdragon 480+ 5G (8 nm)
SM-S906E	n1, n2, n3, n5, n7, n8, n12, n20, n25, n28, n38, n40, n41, n66, n77, n78	Qualcomm Adreno 730 (818MHz)

Table 9: Frequency bands detected in test case scenarios per operator.

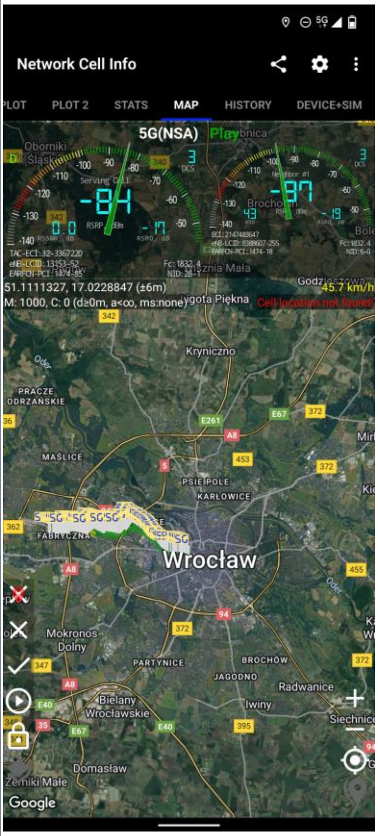
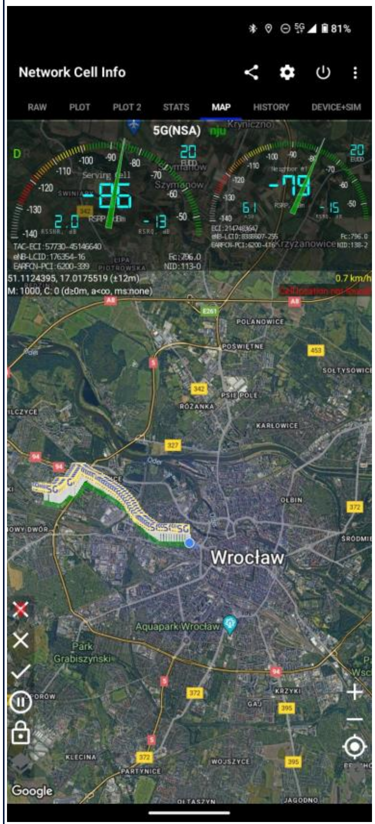
	Urban	Sub-urban	Highway
MNO1	n20, n3, n7/n41, n1	n20, n3, n7/n41, (n1)	n3, (n20)
MNO2	n20, n3, n1	n20	n/a
MNO3	n28, n78	n/a	n/a
MNO4	‘E-UTRA’	n/a	n/a

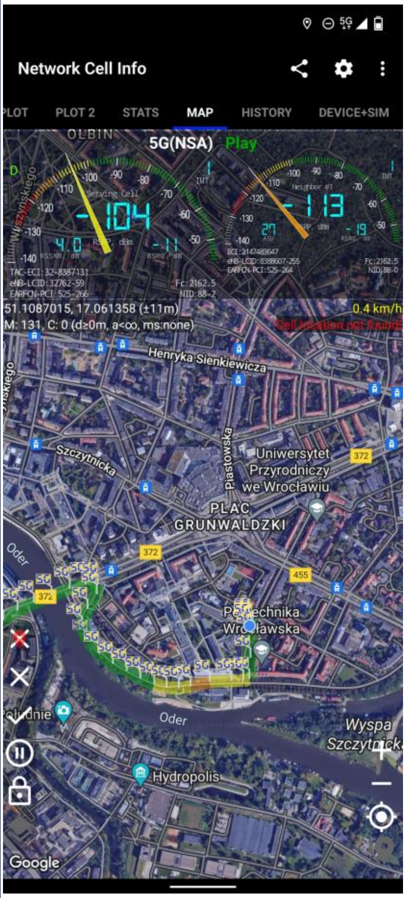
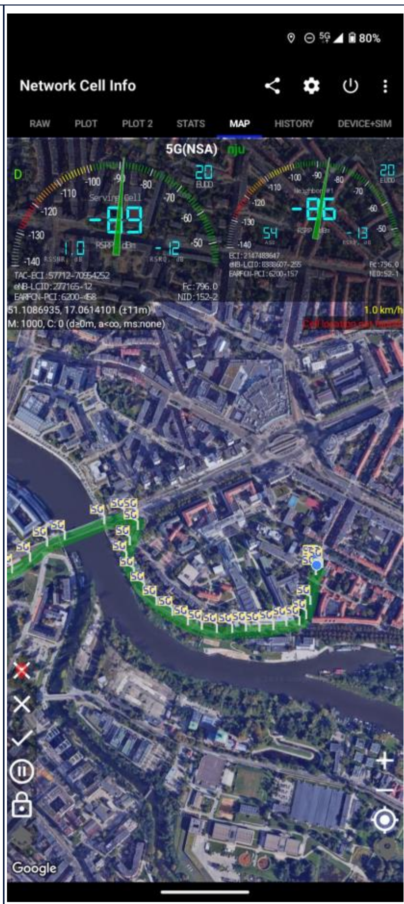
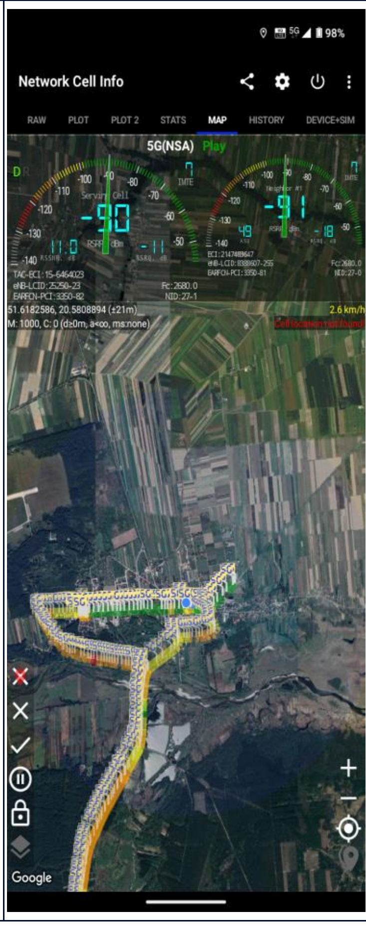
4.2.2 Drive Test Routes

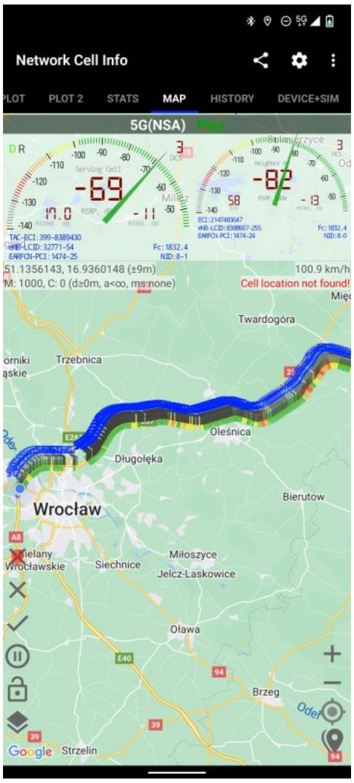
Table 10 shows examples of the displayed information in the ‘Network Cell Info’ application used as a monitoring tool in most of the campaigns. The presented examples show ‘Map’ tab active, that visualized active route and dynamically updated measurements results taken by the device for the serving and neighbor measurements.

Table 10: Exemplary device’s desktop screens, with running ‘Network Cell Info’ Android application during drive test.

Drive test:	Nokia8.3	NokiaG42 5G
a (test start)		

a (mid-test)	Network Cell Info PLOT 2 STATS MAP HISTORY DEVICE+SIM 5G(NSA) 51.1111327, 17.0228847 (±6m) 45.7 km/h M: 1000, C: 0 (d≥0m, a<∞, ms:none) Cell location not found Kryniczno Wrocław Google	81% Network Cell Info RAW PLOT PLOT 2 STATS MAP HISTORY DEVICE+SIM 5G(NSA) 51.1124395, 17.0175519 (±12m) 0.7 km/h M: 1000, C: 0 (d≥0m, a<∞, ms:none) Wrocław Aquapark Wrocław Google

<table>
<tr>
<td>a (test end)</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>b (mid-test)</td>
<td>
</td>
<td>
</td>
</tr>
</table>

c		n/a
d	n/a	n/a

4.2.3 Data Samples

During the drive tests, the application used for network performance monitoring was actively and continuously logging the data, capturing all defined parameters along the planned routes. The samples created comprehensive datasets, with an example shown in : Exemplary data logs.

Table 11: Exemplary data logs.

sim	radiotype	mcc	mnc	cellid	lat	lon	RSRP	RSRQ	SINR	speed	device
1	5G (NSA)	260	6	8399732	51.11504	16.92333	-108	-15	1	-2000	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11488	16.92335	-108	-15	1	5	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11497	16.92326	-107	-15	1	1.5	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11504	16.92323	-109	-16	-1	1.6	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11506	16.92321	-109	-16	-1	0.5	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11506	16.92322	-109	-16	-1	0	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11507	16.92322	-109	-16	-1	0.3	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11509	16.92322	-109	-16	-1	0.5	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.1151	16.92321	-109	-16	-1	0.3	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11511	16.92321	-109	-16	-1	0.1	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.1151	16.92322	-109	-16	-1	0.2	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.1151	16.92322	-109	-16	-1	0	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.1151	16.92322	-109	-16	-1	0.1	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.1151	16.92322	-109	-16	-1	0	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11509	16.92322	-104	-12	4	0.1	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11509	16.92323	-105	-12	4	0.1	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11509	16.92323	-108	-12	4	0.1	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11508	16.92323	-107	-12	8	0	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11508	16.92323	-109	-12	8	0.1	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11508	16.92323	-108	-16	8	0	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11499	16.92333	-108	-16	8	0	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.115	16.92332	-108	-16	8	0	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11499	16.92333	-107	-18	-6	0	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11508	16.92324	-107	-18	-6	0.2	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11508	16.92324	-107	-18	-6	0	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11507	16.92324	-107	-18	-6	0	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11507	16.92324	-107	-18	-6	0	HMD_Global_Nokia_8.3_5G
1	5G (NSA)	260	6	8399732	51.11507	16.92324	-107	-18	-6	0	HMD_Global_Nokia_8.3_5G

4.2.4 Data Limitations

The three basic key performance indicators (RSRP, RSRQ, SINR) collected in the conducted experiment allow for investigations on the RAN performance under any deployment scenario (urban, sub-urban, rural). The experiment inspected eMBB traffic, assuming potential applicability of the resulting findings and framework for the set of devices designed for URLLC and mMTC. However, this is based on theoretical space only. Although, test d delivered user-centric KPIs (throughput and latency), with an attempt to address KPIs that are supposed to differentiate URLLC and eMBB traffic, the data is incomparable. Hence, the findings applicability to any kind of 5G traffic (eMBB, URLLC, mMTC) is weighted with deficiencies.

For the available eMBB inspection from commercial smart phones, there is lack of detailed information on chipsets capabilities and physical layer specific details on how the measurements are derived (including number of antennas' ports, number of Rx, and Tx antennas in use, modulation scheme used). There can be different implementation strategies, thus it may be also limiting factor that make a stand on observed differences or inaccuracies in data collection.

In addition, tests and analyses conducted on 5G performance in NSA deployment in the Frequency Range 1 (FR1) are generally not fully representative of what could be expected in Frequency Range 2 (FR2). This is due to fundamental differences between these two frequency ranges, which affect the performance characteristics and behavior of 5G networks.

4.3 Data Analysis

4.3.1 Trends in parameters

In 5G, the network Reference Signal (RS) are detectable and measurable by the UE. The UE measurements defined in 3GPP specifications, as presented in Table 3. The three RS measurements: RSRP, RSRQ and SINR, that are technically recorded measurement quantities in the drive tests performed under this work, are bound to certain expectations. Each parameter determines the property of strength and quality of the 5G signal and can help to read detailed characteristics of the connection.

The UE measurement of Reference Signal Received (RSRP) is measured for each detectable reference signal as the linear average over the power contributions of the resource elements that carry cell-specific reference signals within the considered measurement frequency bandwidth. In order to make this information usable for the purpose of the analysis, it can be assumed this is the key indicator for signal strength and assessment of RAN performance in context of any radio optimization use case: coverage, mobility and QoS verification and any traffic type (eMBB, URLLC and mMTC). Next to signal strength is signal quality, derived by Reference Signal Received Quality (RSRQ) and Signal-to-Interference-plus-Noise Ratio (SINR).

Reference Signal Received Quality (RSRQ) as defined in Table 3 can be denoted as:

$$RSRQ = N\frac{SS\text{-}RSRP}{RSSI} \tag{1}$$

where *SS-RSRP* is signal power of synchronization signal, *N* is the number of Resource Blocks, Received Signal Strength Indicator (RSSI), comprises the linear average of the total received power observed only in certain OFDM symbols of measurement time resources, in the measurement bandwidth from all sources the UE can see, including co-channel serving and non-serving cells, adjacent channel interference, thermal noise etc. The measure is derived value from RSRP, and it is supposed to give some information about interference in addition to the strength of the wanted signal.

Signal to Interference plus Noise Ratio (SINR), is the measurement for the quality of the signal by comparing the signal strength to the level of interference and noise, which is denoted as:

$$SINR = \frac{S}{I+N} \tag{2}$$

where S is the signal power of determined synchronization and reference signals, I is the interference power distribution, and N is the noise power distribution. Higher SINR values indicate positive result of the ratio and stronger network signal over the linear average of the noise and interference power contribution. It reveals some commonality with RSRP, through reference to interferences.

Each parameter that determines the quality of the 5G signal can help to read detailed characteristics of the connection (Table 12). RSRP as the average power of the cell-specific reference signals implies the level of the received signal from the base station. It is commonly understood the higher values, the better. For instance, with RSRP = -80dBm, the connection is perceived as good, while with RSRP= -120 dBm and below, the connection may be perceived as poor. RSRQ range considered as poor falls into -20dB and below. It indicates typically high cell load or interference level. At SINR values below 0, the connection perception will be poor (e.g., very slow connection), since this value characterize the situation when there is more noise in the received signal than the useful part, and the probability of losing a connection is high [92]. Yet, it may reflect a relative distance from base station, that varies in dependence on used frequency range (Figure 21: Signal range dependency on frequencies.).

Table 12: Radio performance indicators value ranges.

Radio conditions	RSRP (dBm)	RSRQ (dB)	SINR (dB)
Excellent	>= -80	>= -10	>= 20
Good	-80 to -90	-10 to -15	13 to 20
Medium	-90 to -100	-15 to -20	0 to 13
Poor	<= -100	< -20	<= 0

The generalized overview may suggest that the key indicators are related: the higher RSRP, RSRQ and SINR the better signal quality. Although, due to the various factors that influence the indicators differently, the relationship between RSRP, RSRQ and SINR requires more insightful inspection.

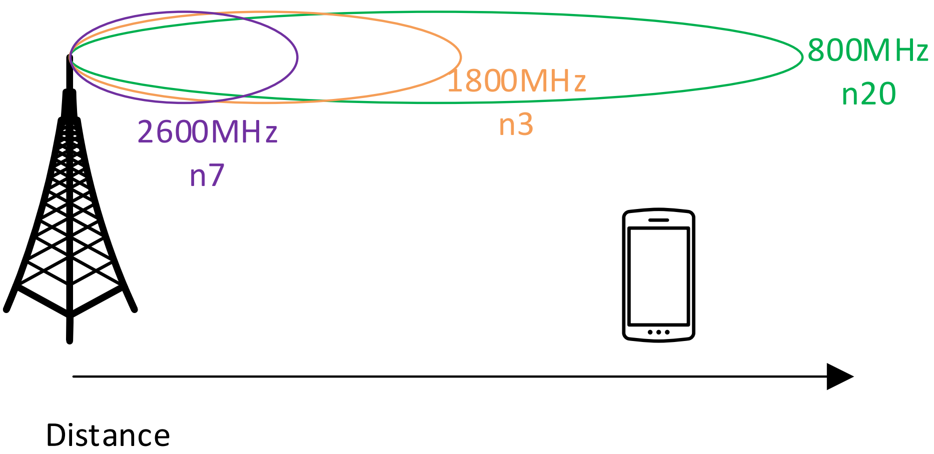


Figure 21: Signal range dependency on frequencies.

4.3.2 Results analysis

This section analyses the collected data in different test case scenarios, presents the results and the results analysis. The collected data from each test case scenario is presented with RSRP, RSRQ, SINR distributions and plot pairs, in case test were simultaneously performed by two devices. Histograms are used to display the modality in data distribution and parameters values concertation, insights into the frequency, spread and central tendency for each of the parameter per deployment option. Alternate form of distribution is presented in vertical bar charts, for convenient interpretation of data granularity, their minimum, maximum and median value (marked with red line) and immediate spot of differences between parameters of the same category but recorded at different devices or test case scenario.

4.3.2.1 Results for Urban area

This section provides visualization of the collected data from urban areas. Figures 22-26 present the distribution of data collected in test case a, performed in urban area scenario tested by two devices simultaneously, each in service of different MNO. Figures 27-28 presents outcome of data collection in the test case d, with urban area scenario tested by one device.

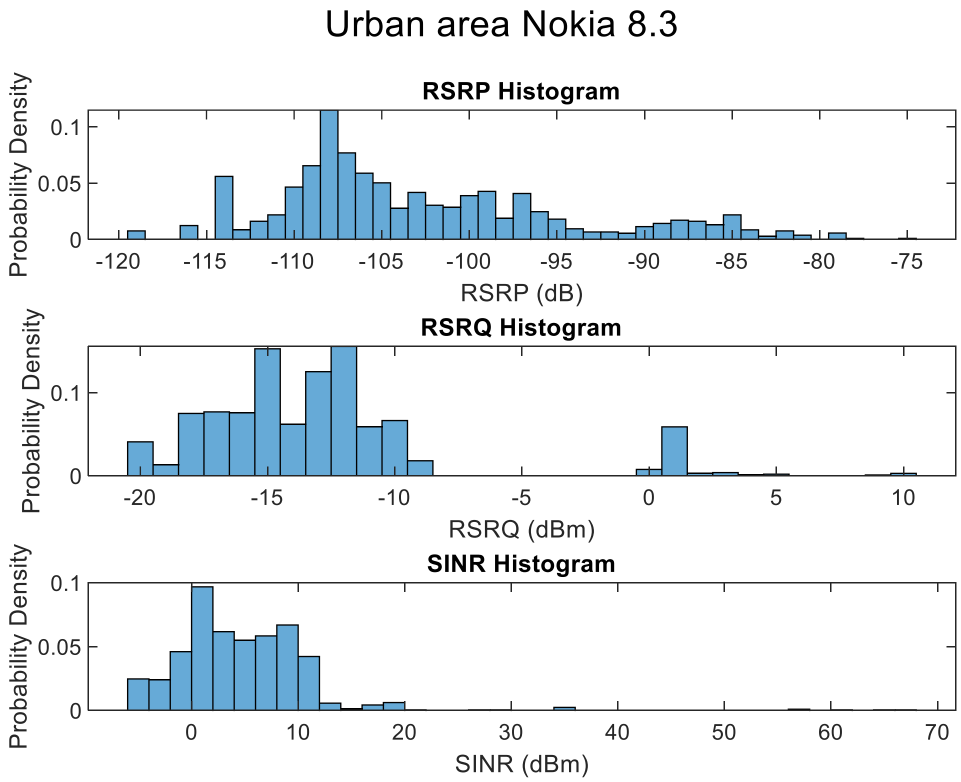


Figure 22: Distribution histogram of RSRP, RSRQ, SINR in urban area for MNO1.

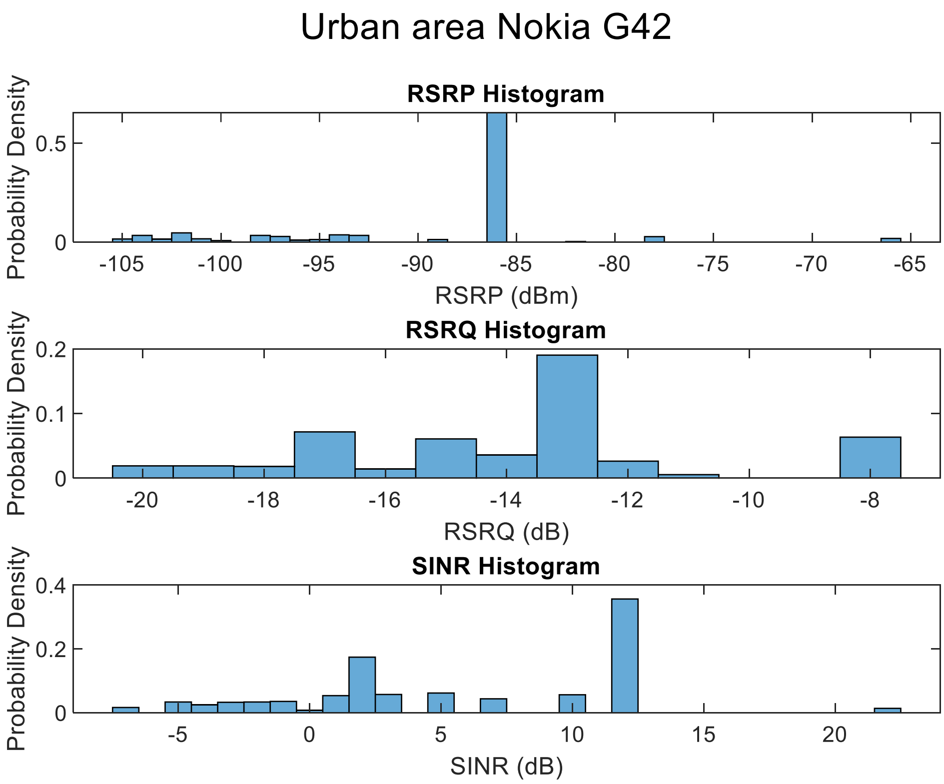


Figure 23: Distribution of RSRP, RSRQ, SINR in urban area for MNO2.

Figure 22 Figure **22**: Distribution histogram of RSRP, RSRQ, SINR in urban area for MNO1.and Figure 23 present data collection from the same area. There are notable differences in signals distribution. Due to the same area and topography, the differences likely result from various strategies on network deployments among the two tested operators and devices properties. As the utilized commercial devices are equipped with chipsets that do not differ notably with offered capabilities [93][94], it remains inadequate how much devices' chipsets impact the observed differences in results. However, one potential understanding can result from differences in frequency bands support. Conversely, the distinctions can be deduced from different networks setup by the two MNOs. In the MNO1 network, there were more frequent changes of serving cells observed with higher probability to perceive signal degradation due to higher frequencies range deployed, frequent switching between the cells, and therefore, likely more neighboring cells interferences experienced. Better signals propagation in MNO2, suggest that the deployment offers more stable connection, with better signals perception (due to longer stays in low-band frequency). The RSRP histogram presents one distinct pick, which relates to the stable distribution. It is not the same for RSRQ and SINR, teaching that signal quality was impacted and changing more. As shown in Figures 24-26, all the parameters median values for RSRP, RSRQ and SINR were higher in MNO2, comparing to MNO1. Also 'taller' bar charts for RSRQ and SINR reveal widely varying ranges of parameters related to signal quality in MNO2 network.

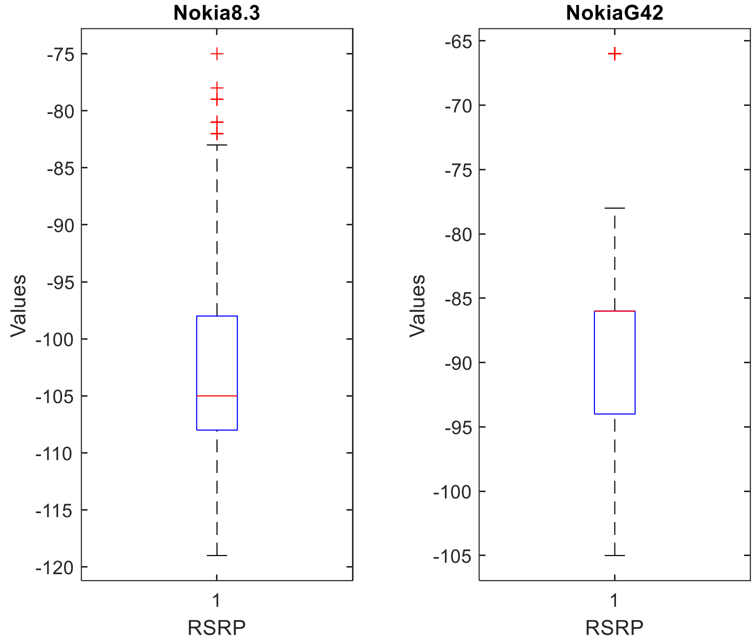


Figure 24: Bar graph for distribution of RSRP in urban area for MNO1 (Nokia8.3) and MNO2 (NokiaG42).

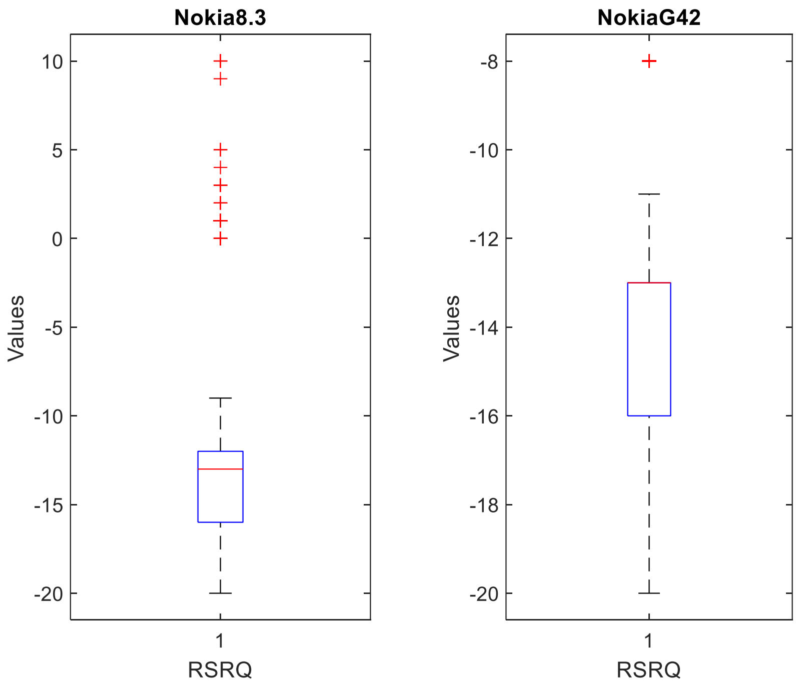


Figure 25: Bar graph for distribution of SINR in urban area for MNO1 (Nokia8.3) and MNO2 (NokiaG42).

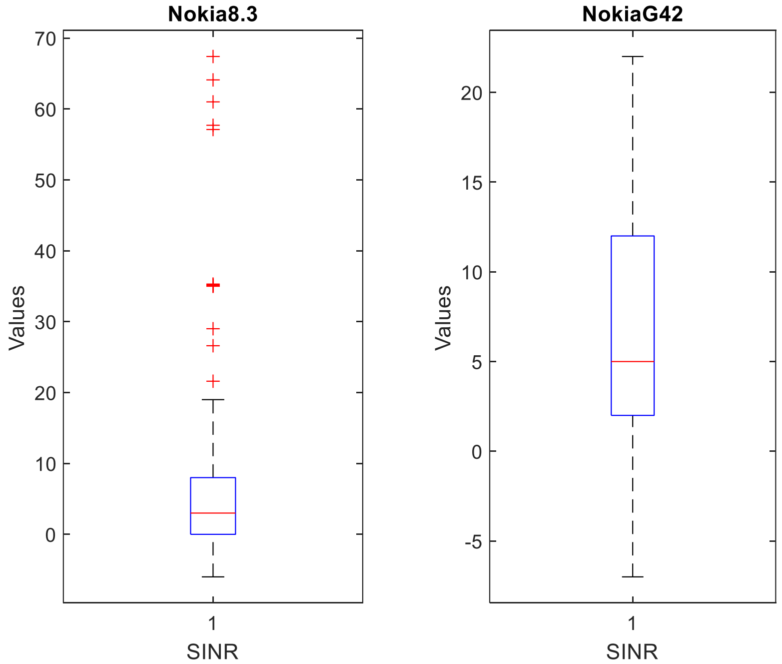


Figure 26: Bar graph distribution of SINR in urban area for MNO1 (Nokia8.3) and MNO2 (NokiaG42).

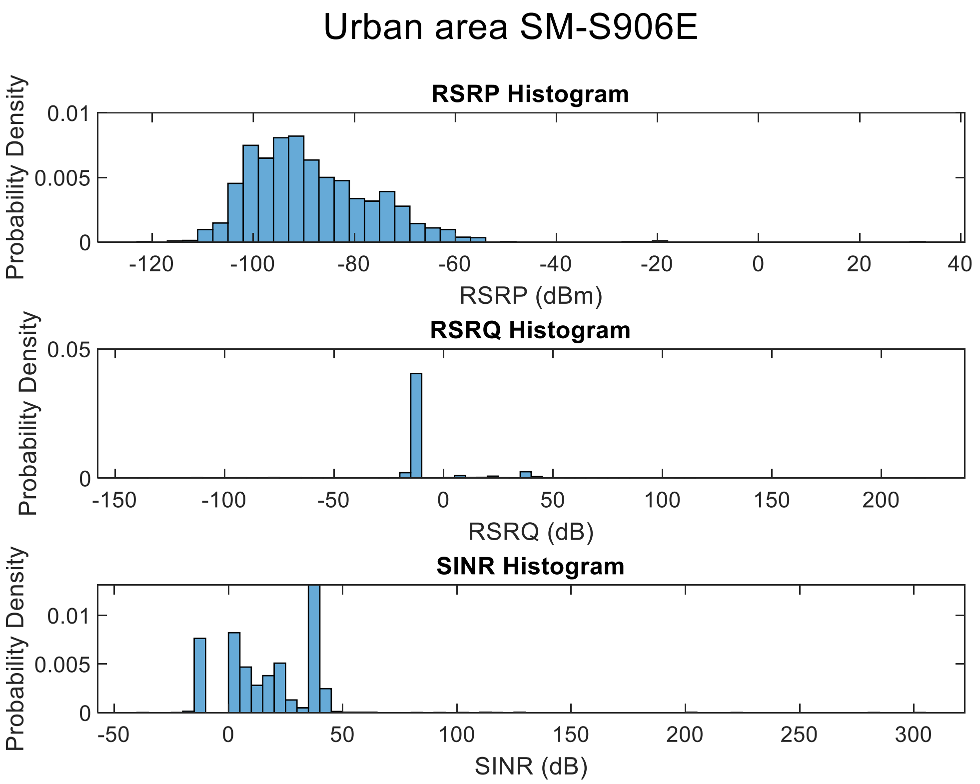


Figure 27: Distribution of RSRP, RSRQ, SINR in urban area for MNO3.

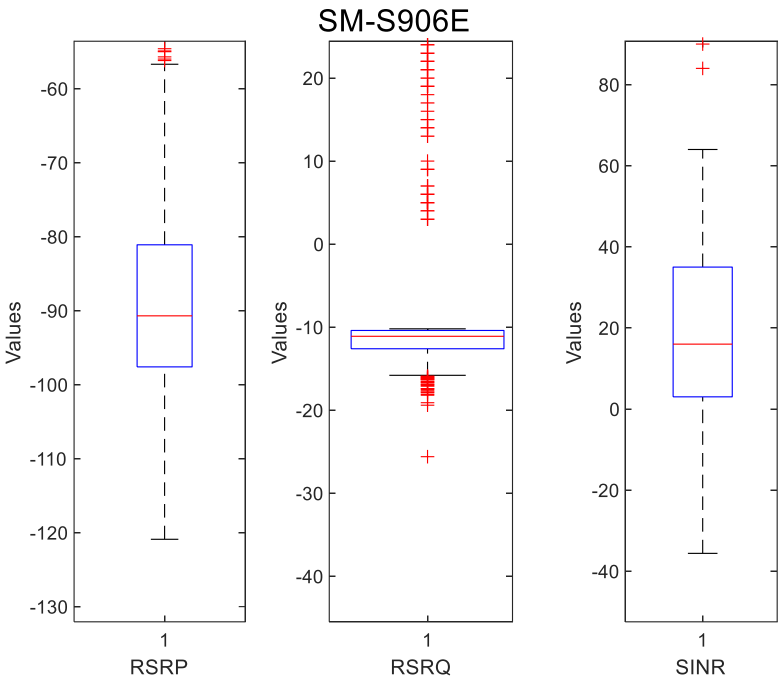


Figure 28: Bar graph for distribution of RSRP, RSRQ, SINR in urban area for MNO3.

For comparison with different urban area, Figure 27 depicts a distribution of RSRP, RSRQ, SINR in deployment using two frequency ranges (n28 and n78) (test case d). Although higher frequency was in use, range of the recorded metrics' results in this data collection displays higher median values than results for MNO1. It may be a consequence of only one (in contrary to MNO1) higher frequency range and TDD technique in use, which should enable relatively static network setup.

4.3.2.2 Results for Sub-urban and Rural area

This section provides visualization of the collected data from sub-urban area. Figures 29-33 present the distribution of data collected in test case b, performed in sub-urban area scenario tested by two devices simultaneously, each in service of the same MNO (also the same operator as in test case a denoted as MNO1).

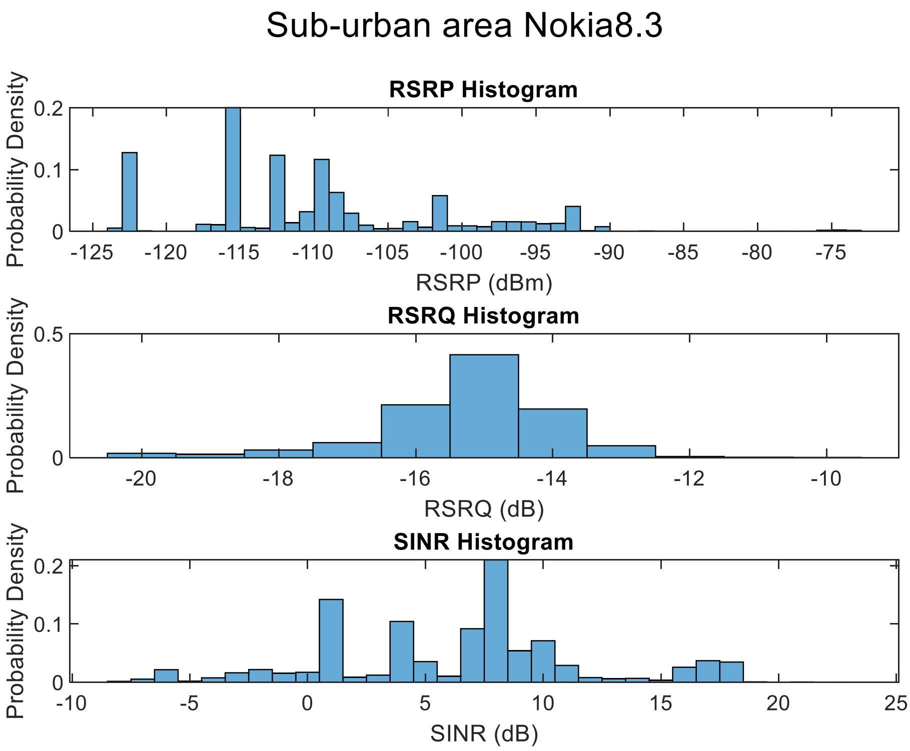


Figure 29: Distribution of RSRP, RSRQ, SINR in sub-urban area for MNO1 (Nokia8.3).

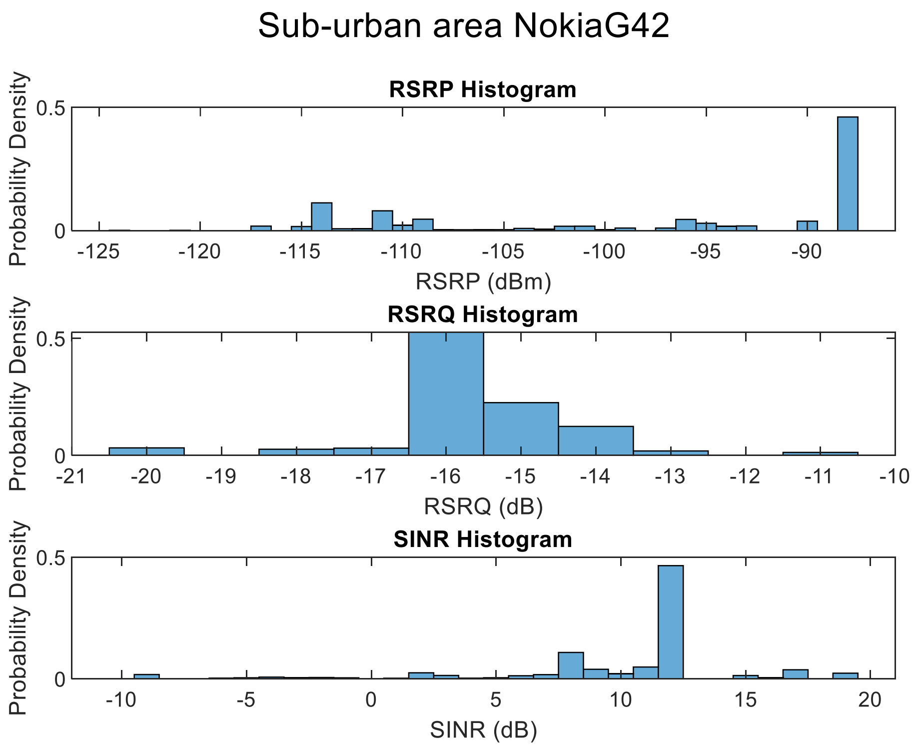


Figure 30: Distribution of RSRP, RSRQ, SINR in sub-urban area for MNO1 (NokiaG42).

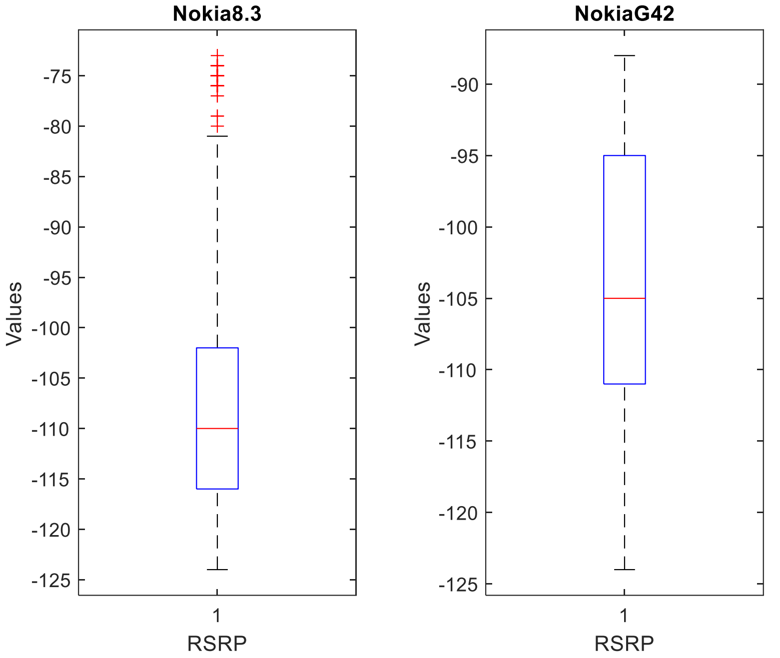


Figure 31: Bar graph for distribution of RSRP in sub-urban area for MNO1 (Nokia8.3, NokiaG42).

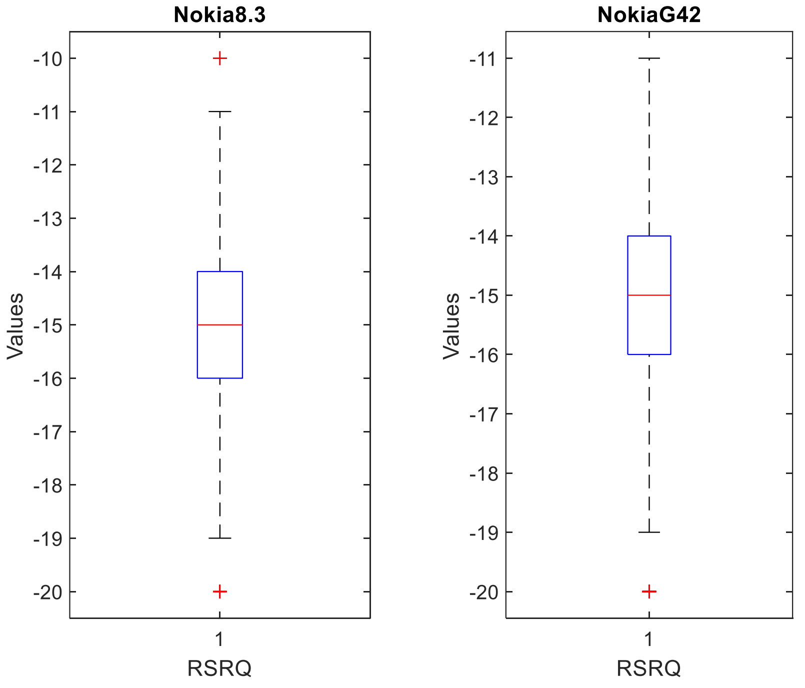


Figure 32: Bar graph for distribution of RSRQ in sub-urban area for MNO1 (Nokia8.3, NokiaG42).

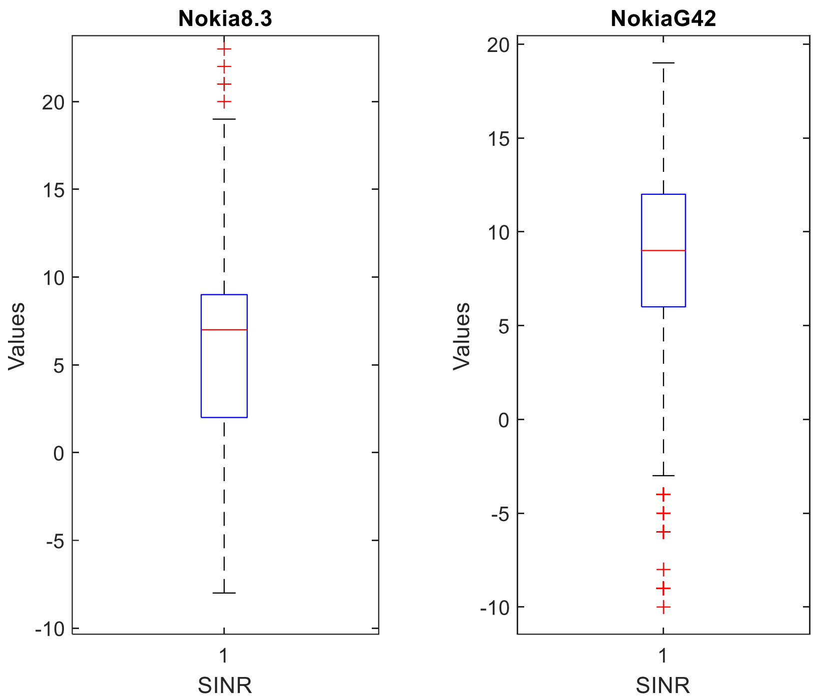


Figure 33: Bar graph for distribution of SINR in sub-urban area for MNO1 (Nokia8.3, NokiaG42).

For the same route and the same operator tested, the signals distributions reveal similarities and very comparable results mainly for RSRQ. However, RSRP and SINR present some differences in distribution and maximum values achieved during the test. To make it explainable Figure 34 shows one relevant recorded situation during the test. It concerned 5G coverage edge, according to Nokia8.3 display (left hand side in the Figure 34), and changing the carrier to 4G. According to NokiaG42 displayed information, detected signal was still 5G and at an acceptable level. In this particular interval, there was noticeable difference in RSRP signal value: one device measured RSRP at the level of -123dBm, and the other at -114dBm, whereas RSRQ and SINR were at comparable levels. The behaviour can hint towards higher sensitivity of the device to switch to another detectable carrier, which can be acknowledged from the wider range of frequencies supported by Nokia8.3 (Table 8). The longer bar for RSRP presented for Nokia G42, can be considered as a corresponding and suitable result with the impact brought by the coverage edge situation. The height of the bar is proportional to the wide range of values it represents, including strong signals and the weakest. Due to fewer changes of the serving cell/carrier, the bar chart for Nokia G42 RSRP represents a wider range of RSRP values of the 5G carrier.

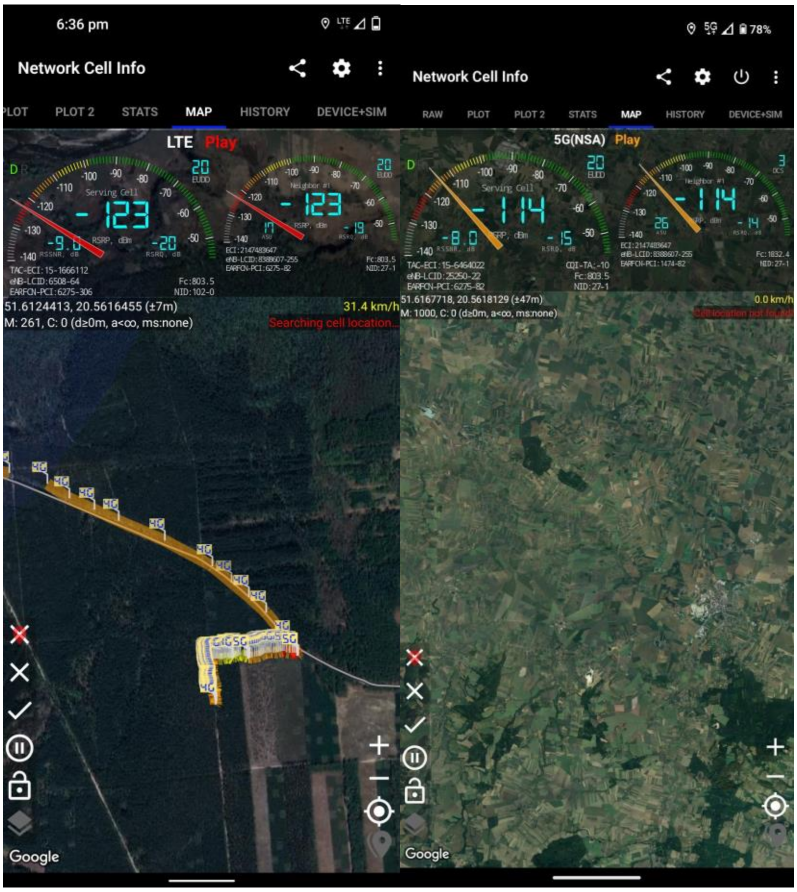


Figure 34: Loss of 5G signal in sub-urban area by Nokia8.3.

4.3.2.3 Results for Highway area

This section provides visualization of the collected data from a highway road. Figures 35-36 present the distribution of data collected in this test (case d), performed with one device, in a service of one MNO (denoted previously, in different test cases as MNO1).

The results for this test scenario present varying picks in histograms, with the most distinct pick for RSRP value: -110dBm, RSRQ: -15dB and SINR: 8 dB. Nonetheless, its worth noting this drive test recorded the highest RSRP values among all the conducted drive test (with picks at -64dBm). This (together with SINR positive pick value) can suggest good signals perceptions from the frequencies along highway, which are usually transmitted with close proximity of the base stations to the road. Perception of the strong signals appears to taper any significance of the car velocity, which didn't impact RSRP and SINR results in a negative manner. RSRQ results do not record a significant variability in the data, but remain at low level.

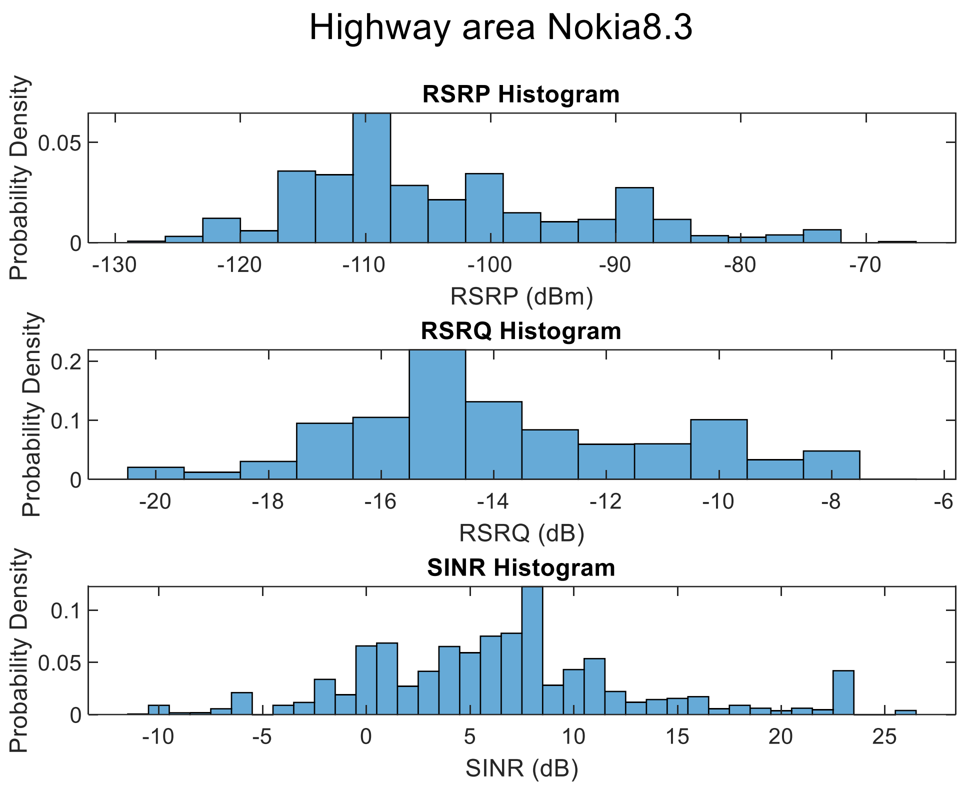


Figure 35: Distribution of RSRP, RSRQ, SINR in a highway area for MNO1.

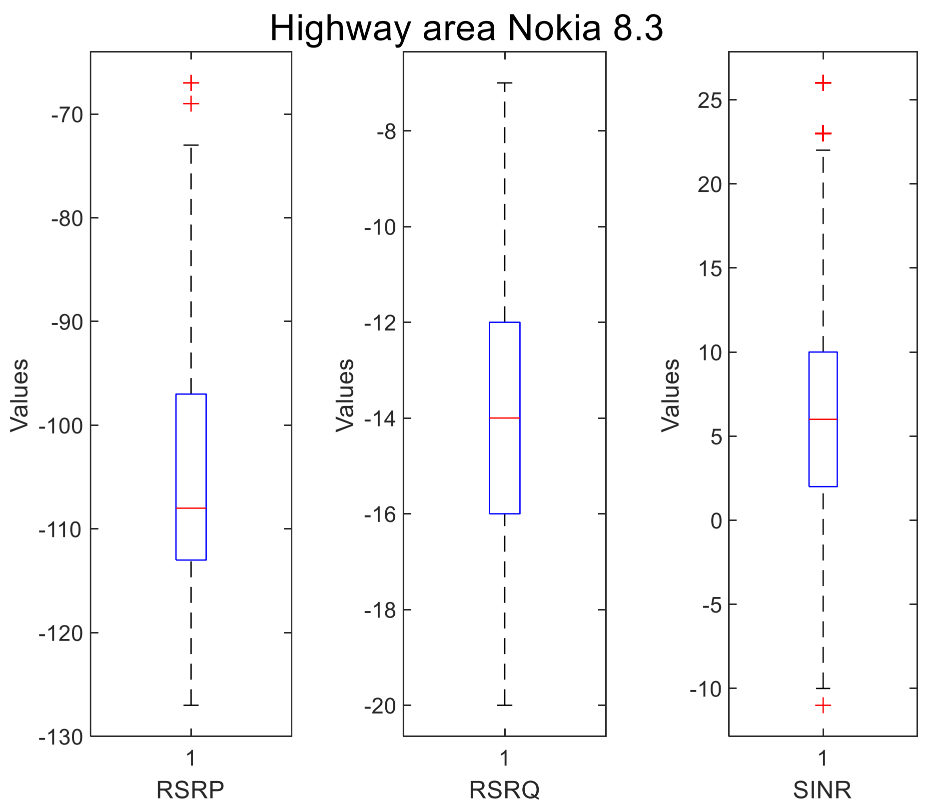


Figure 36: Bar graph for distribution of RSRP, RSRQ, SINR in a highway area for MNO1.

4.3.2.4 Drive Test results summary

In most cases the KPIs do not reveal symmetrical distribution, therefore the data are collectively analyzed in this subclause with reference to the median values. This enables to present central location of the data due to better suitability for measure of central tendency for skewed distributions. The expression of the central tendencies for the tested radio parameters per use case are presented in Table 13:

Table 13: Measurements results from the conducted test in different scenarios.

KPI		Urban			Urban (MNO3)			Sub-urban/Rural			Highway		
		Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max
RSRP	MNO1	-105	-119	-75	-104	-120	-50	-107	-124	-73	-105	-127	-67
	MNO2	-86	-105	-66				-93	-124	-88			
RSRQ	MNO1	-15	-20	-10	-11	-12	-10	-15	-20	-10	-14	-20	-7
	MNO2	-13	-20	-8				-16	-20	-11			
SINR	MNO1	3	-6	32	16	3	35	7	-8	23	6	-11	26
	MNO2	5	-7	22				12	-10	19			

Based on the thoroughgoing analyses, findings on the RSRP, RSRQ and SINR relationships in the tested scenarios, are the following:

- RSRP: values typically range from around -120 dBm (very poor) to -60 dBm (excellent). Highest values have been recorded in urban and highway area (-66dBm and 67dBm, respectively). Highest median for RSRP has been recognized in urban scenario. Best distribution (with data symmetrically distributed around central mean) has been observed in urban area operating on lower frequencies, displaying high durability and stability of the parameter.
- RSRQ: values typically range from -20 dB (very poor quality) to -10 dB (good quality). Highest values have been recorded in urban area operating at higher frequencies. Highest median for RSRQ has been recognized in rural scenario. Best distribution has been observed in rural area, presenting low variability of the parameter.
- SINR: values typically range from -8 dB (very poor quality) to 20 dB (good quality). Highest values have been recorded in urban area. Highest median for SINR has been noted in rural scenario There seem to be no appealing, unimodal distribution observed in any of the test scenarios, revealing the parameter variability and impulsiveness to the environment, with dynamic changes from high to low value and having wide distributions, especially in urban area.

RSRQ with its unimodal distribution in sub-urban area (Figure 29: Distribution of RSRP, RSRQ, SINR in sub-urban area for MNO1 (Nokia8.3).) suggests it should be easy to achieve comparable results among different users. However, its sensitivity changes depending on the scenario and there was no comparable distribution observed in other scenarios (urban vs. sub-urban/rural). Similarly, SINR, which indicates slightly better trends in sub-urban scenario, appears to be more sensitive factor (compared to RSRQ) in changes of sub-urban area to rural (Figure 33). Since RSRQ and SINR can change dynamically and more frequently, being sensitive to terrain, physical constrains, trees, obstacles and environment, RSRP asset appears to result from lower sensitiveness to the environmental changes. The RSRP became also the final determinant at the coverage edge situation (in rural area), helping to determine actual signal perception among the two devices. RSRP seems most diagnostical across different distributions. Though its distribution varies, making it challenging to determine the parameter value with confidence, hence additional investigations on the collected data correlation are conducted in the next section to draw more definitive conclusions.

4.3.3 Trends in correlation

A correlation technique is explored to determine if any better statistical method can prove further relationships among the collected data. The correlation-based methodology typically used to determine the degree to which two variables are related.

RSRP, RSRQ, SINR parameters qualifies to apply correlation coefficient measure due to normalized meaning of the parameters. Table 10 displays various parameters values that relate to different degrees of radio network perception. If a correlation coefficient for RSRP, RSRQ, SINR can result in a proof that the performance indicators in some scenarios depend on each other [55], the observations on data collection trends conducted in sub-clause 4.3.3 can get more conclusive. Consequently, conducting this research with the ultimate goal to justify the data selection process for the target ML-based methodology, could get evidence on the data dependency, and justification to possibly reduce the data dimension in related Data Collection.

The correlation coefficient is a statistical measure that describes the strength and direction of a relationship between two variables. In its broadest definition, refers to a metric that quantifies the relationship between variables. In data that are correlated, a change in the size of one variable corresponds to a change in the size of another variable, which can occur in the same (positive correlation) or in the opposite (negative correlation) direction. Typically, the concept of correlation is applied to describe a relationship between two continuous variables and is represented by the Pearson product-moment correlation. The Pearson correlation coefficient is generally applicable to data that are jointly normally distributed (following a bivariate normal distribution). For continuous data that do not follow a normal distribution, for ordinal data, or for datasets that contain significant outliers, a Spearman rank correlation can be utilized to assess a monotonic relationship. Both types of correlation coefficients are standardized to range from -1 to +1, where a value of 0 signifies no linear or monotonic connection, and the relationship intensifies, tending toward a straight line (in the case of Pearson) or a consistently increasing or decreasing curve (in the case of Spearman) as the coefficient nears an absolute value of 1 [95]. The correlation coefficient ranges from -1 to 1, represented by an ‘r’ value:

$$-1 \leq r \ \leq 1 \text{ , where:} \tag{3}$$

- r =1 implies perfect positive correlation. As one variable increases, the other also increases in a perfectly linear relationship.
- $r = -1$ implies perfect negative correlation. As one variable increases, the other decreases in a perfectly linear relationship.
- r = 0 indicates no correlation. There is no linear relationship between the variables.

The absolute value of the coefficient (its size, ignoring the sign) indicates how strong the relationship is between two variables. The closer the value to zero, the weaker relationship is. The closer the value to 1, the stronger relationship is.

4.3.4 Correlation analysis

This section examines possible relationships among 5G measurement parameters collected as experimental data (RSRP, RSRQ, SINR), based on evaluating their relation degree and strength with the statistical correlation technique. The correlations for each pair out of the collected parameters: RSRP, RSRQ and SINR in the two test case scenarios: a (urban) and b (sub-urban), where two devices simultaneously recorded data, to test two sets of parameters inputs for comparison. Furthermore, correlation has been used to check relationship between throughput and latency with the key radio parameters (RSRP, RSRQ, SINR) for the test case c, where the two additional attributes were available.

The computation of cross-correlation among each pair of the parameters from the collected data set (RSRP, RSRQ, SINR, and throughput and latency for test d) has been performed with *corrcoeff (x,y)* function in MATLAB DSP (Digital Signal Processing) Toolkit. The function is modelled in the environment as depicted in the Figure 37.

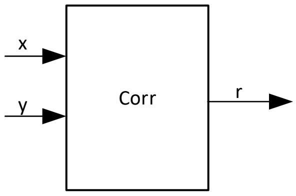


Figure 37: Block model for computing cross-correlation.

The operation in the correlation block can be presented as a correlation coefficient function, using the formula (3), which computes relation of the two input parameters, and calculates an outcome *r* representing relationship strength of the two parameters in time domain:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \tag{4}$$

where:

- n is the number of data points,
- x and y are the individual data points for the two variables,
- $\sum xy$ is the sum of the product of paired scores,
- $\sum x$ and $\sum y$ are the sums of the x and y scores, respectively.

The experiment with applying a correlation coefficient for RSRP, RSRQ, SINR can result in a proof that the performance indicators in some scenarios depend on each other. Consequently, it can justify the data selection process for the target ML-based methodology, and justification to possibly reduce the data dimension in related Data Collection. The computation of cross-correlation among each pair of the parameters from the collected data set (RSRP, RSRQ, SINR) has been performed with *corrcoeff (x,y)* function. The nominal value of correlation for any pair of parameters can be equal 1. If such value if observed, the tested parameters in the pair can be exchangeable in ML algorithm Input Data, and one of these parameters can be omitted (if the other is present). If there is a weak relationship, with numerical value ranging from (0.1-0.35), the tested parameters in the pair needs to be considered in ML algorithm Input Data, and none of these parameters can be omitted detailed observations,

4.3.4.1 Results for Urban area

In this section the visualization of correlation coefficient function applied to RSRP, RSRQ and SINR data collected in urban scenario (test case a) is presented. The area was inspected with two devices simultaneously (Nokia 8.3 and Nokia G42). During the drive test, each device was in service of different MNO, whereas the route was the same.

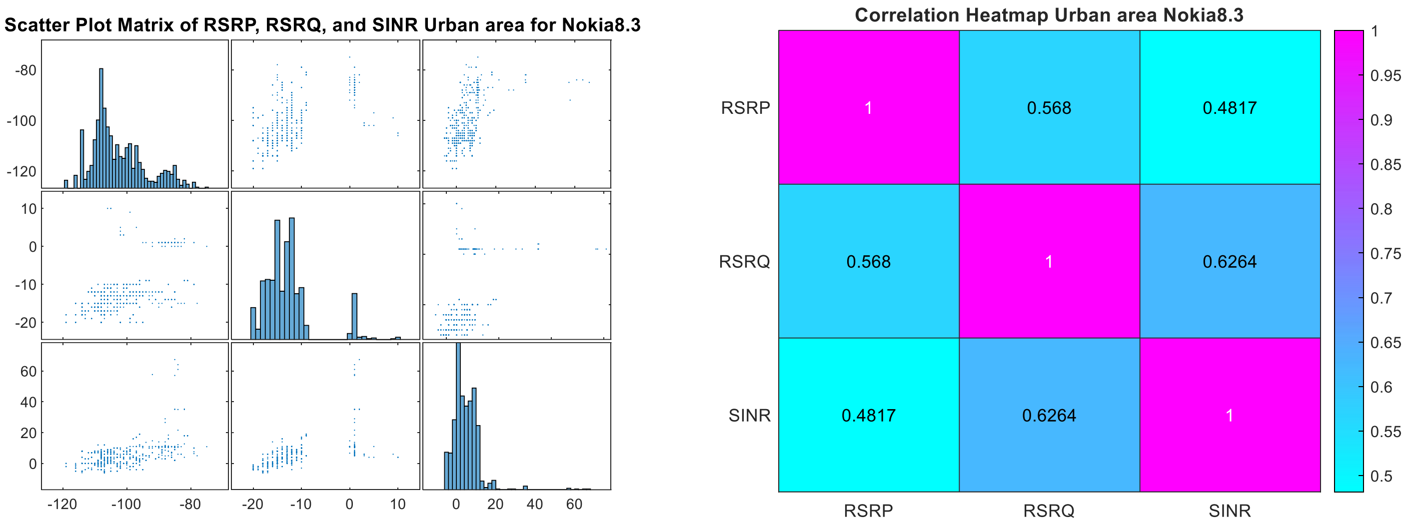


Figure 38: Correlation co-efficient for RSRP, RSRQ, SINR in Urban area, MNO1 (Nokia8.3).

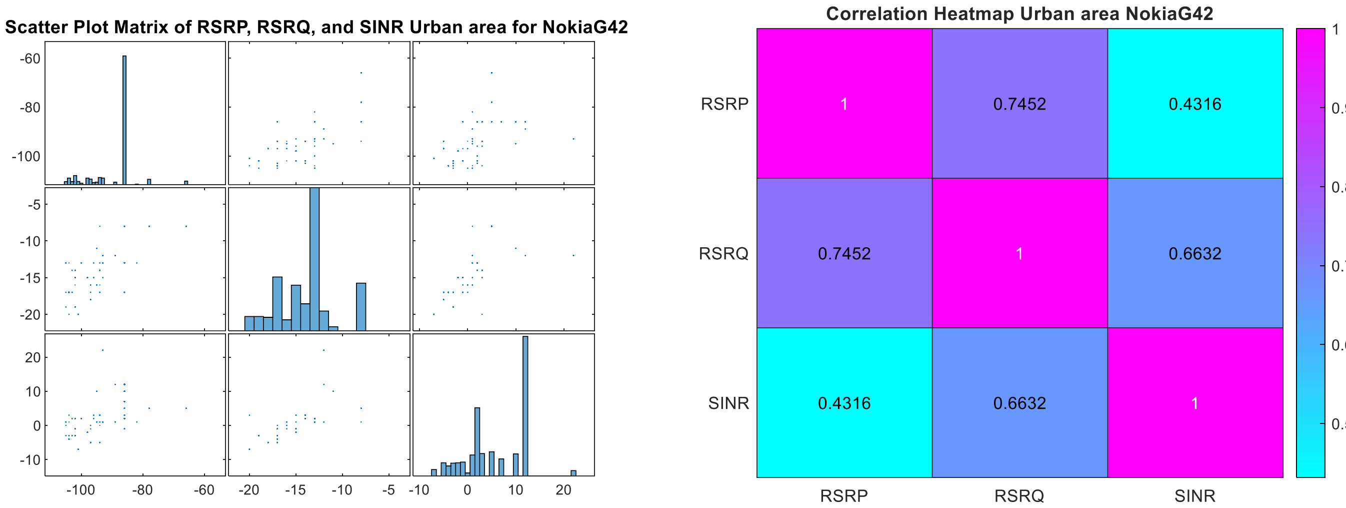


Figure 39: Correlation co-efficient for RSRP, RSRQ, SINR in Urban area, MNO2 (NokiaG42).

Figure 38 presents scatter plot of RSRP, RSRQ and SINR and corresponding table of correlation heatmap with calculated correlation coefficient value *r* for each of the parameters' pair for Nokia 8.3 in MNO1 operator's service.

Figure 39 presents scatter plot of RSRP, RSRQ and SINR and corresponding table of correlation heatmap with calculated correlation coefficient value *r* for each of the parameters pair for Nokia G42 in the service of operator MNO2.

Comparing correlations, it can be observed that in two test case scenarios, the three parameters have positive coefficient. The figures on each pair of the tested data (RSRP and RSRQ, RSRP and SINR, RSRQ and SINR) prompt increasing slopes upward on the scatter plot matrix graphs. It shows the trend that in case of one parameter increase the other is also increasing. The nominal value is observed for RSRP and RSRQ pair correlation for the MNO2 (0.7452), suggesting that in this scenario RSRP and RSRQ correlation between the two parameters was stronger than for any other parameters pair. Including the parameters correlation at the other MNO. This result might be matching findings on different deployment strategy and better signals propagation in MNO2, where

the deployment offered more stable connection (compared to MNO1), with better signals perception (due to longer stays in low-band frequency).
However, in both scenarios, there is a common trend for correlation between two parameters designed for signal quality measure: RSRQ and SINR. The correlation between the two parameters is at a comparable level: 0.6264 for MNO1, and 0.6632 for MNO2.

4.3.4.2 Results for Sub-urban area

This section provides results of correlation coefficient function applied to RSRP, RSRQ and SINR data collected in sub-urban scenario (test case c). The area was inspected simultaneously with two devices (Nokia 8.3 and Nokia G42). During the drive test, each device was in service of the same MNO and the route was the same.

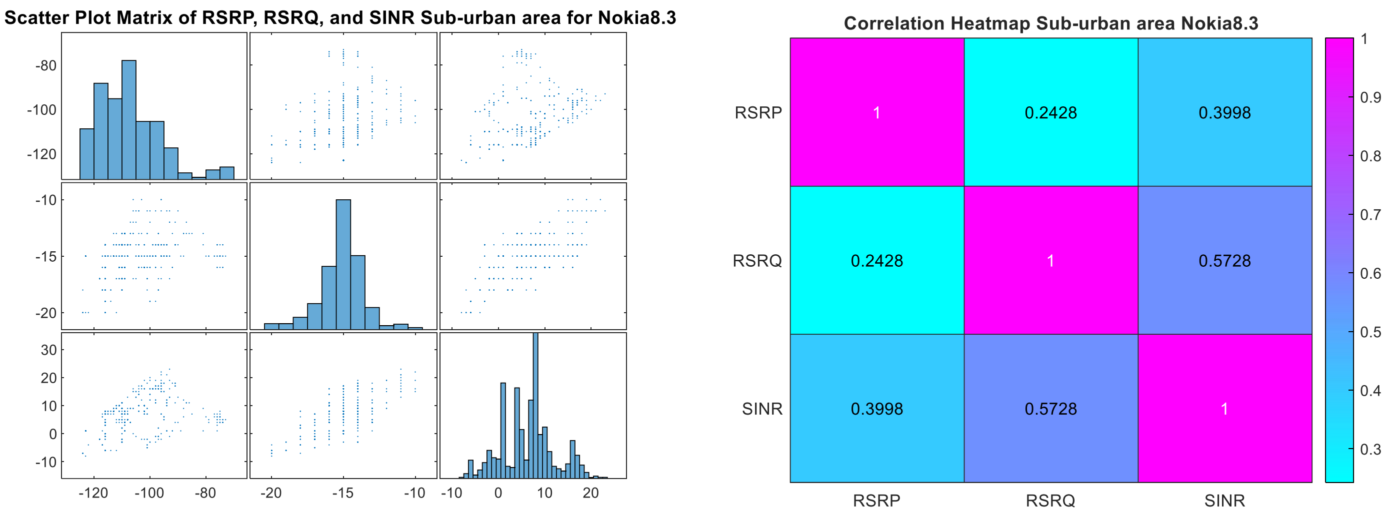


Figure 40: Correlation co-efficient for RSRP, RSRQ, SINR in sub-urban area, MNO1 (Nokia8.3).

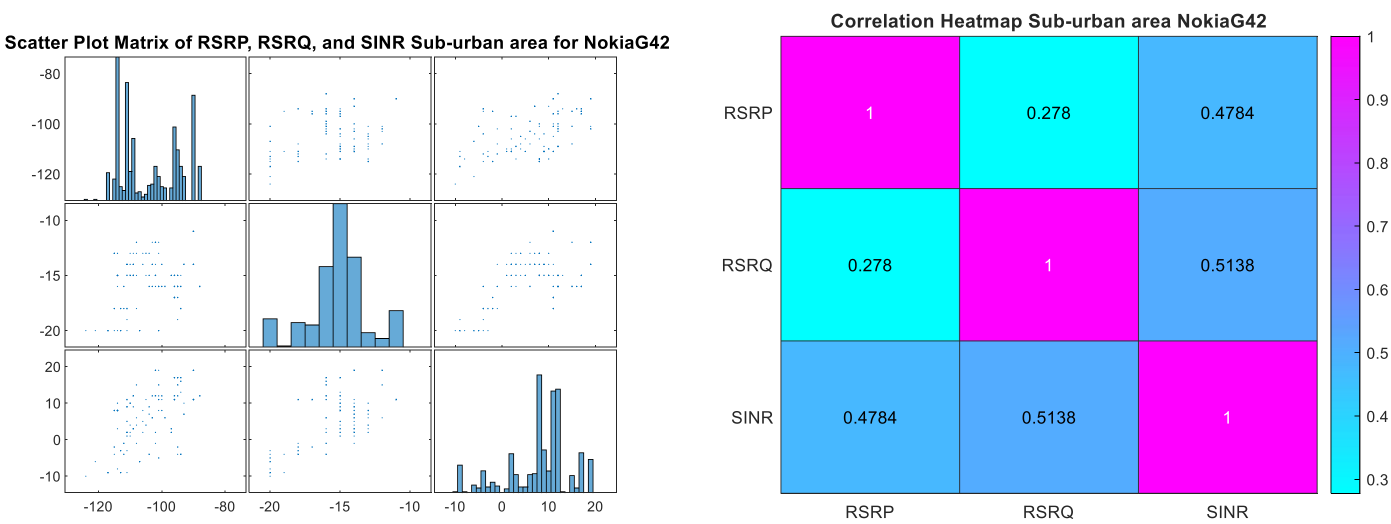


Figure 41: Correlation co-efficient for RSRP, RSRQ, SINR in sub-urban area, MNO1 (NokiaG42).

Figure 40 presents scatter plot of RSRP, RSRQ and SINR and corresponding table of correlation heatmap with calculated correlation coefficient value *r* for each of the parameters' pair for Nokia 8.3 in MNO1 operator's service.

Figure 41 presents scatter plot of RSRP, RSRQ and SINR and corresponding table of correlation heatmap with calculated correlation coefficient value *r* for each of the parameters' pair for Nokia G42 in the service of the same operator MNO1.

Each pair (RSRP and RSRQ, RSRP and SINR, RSRQ and SINR) of the tested data set prompts increasing slopes upward on the scatter plot matrix graphs, which proves positive correlation. For the two devices, the highest correlation is observed for RSRQ and SINR pair. Comparing to urban scenario, there is a remarkable change in correlation of RSRP with other metric, and for two devices the trend is comparable.

4.3.4.1 Results for Urban area: additional parameters

This section provides results of correlation coefficient function applied to KPIs directly related to user performance measure: throughput and latency. 'NR PCell Throughput' and 'Time to First 500Kbytes' were supplementary data collected in urban scenario (test case c) in MNO3. The area was inspected by one device and during one drive test. Assuming these two parameters can be representative for throughput and latency, correlation is applied to check each of the two parameters relation with radio specific KPI (RSRP, RSRQ, SINR).

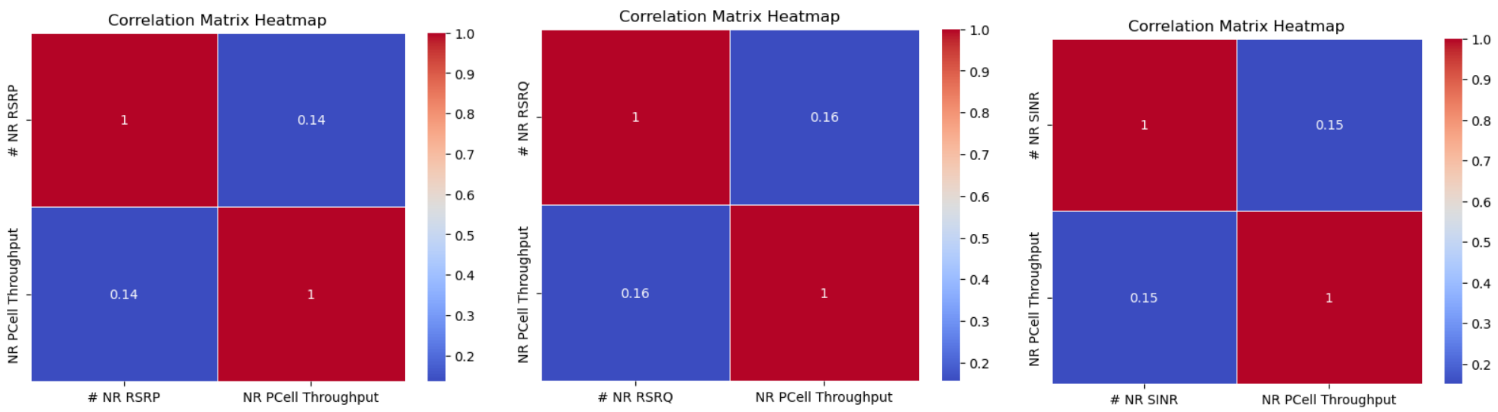


Figure 42: Correlation co-efficient for user's throughput with RSRP, RSRQ and SINR in urban area, MNO3.

The check on correlation co-efficient of user's measured NR PCell Throughput with any of the three radio performance measures (RSRP, RSRP, SINR) resulted in similar values, depicting there weak relationship (0.14 or 0.16). Based on the results (Figure 42), it can be concluded that the changes of RSRP, RSRQ or SINR didn't influence NR PCell Throughput considerably, and there is no appreciable relationship between throughput and key radio KPIs. This can guide there is another dimension, that should be taken into account for assessment of the throughput impacts. The result matches the theory stating that user's achievable throughput is dependent on number of secondary carriers and what the network configures on top of the primary cell (PCell). Conversely, the results prove that for basic and primary carriers the user's throughput impacts are negligible.

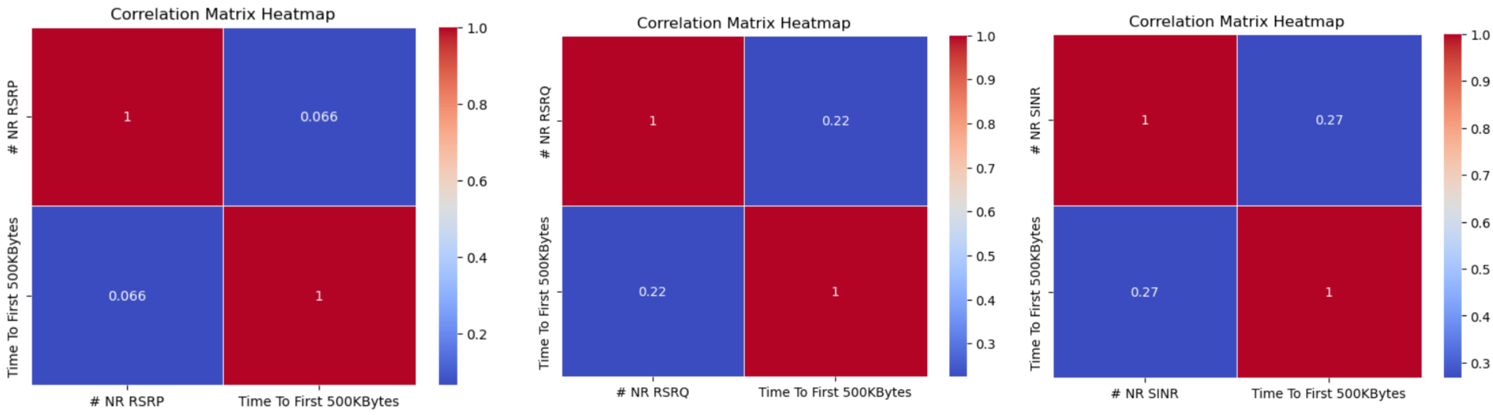


Figure 43: Correlation co-efficient for user's latency with RSRP, RSRQ and SINR in urban area, MNO3.

The check on correlation co-efficient of user's latency with any of the three radio performance measures (RSRP, RSRQ, SINR) (Figure 43), reveals there is stronger relationship between latency and signal quality (RSRQ, SINR) than between latency and signal strength (RSRP). This can guide that for users, whose KPI is strictly determined by latency (URLLC), signal strength shouldn't be the only determinant in performance assessment.

As the measurement results achieved in test case d, could not be repeated, nor compared with drive tests for the other scenarios (due to logging application limitations not recording such parameter), the results are considered as limited. Though, they provide insight into what information can be critical for 5G traffic types tied to stringent latency requirements.

4.3.4.2 Correlation analysis summary

Correlation co-efficient applied to RSRP, RSRQ and SINR, helped to determine relationships among the collected data. Assuming strong, moderate and weak relationship refer to correlation co-efficient numerical value ranges: (0.75-1), (0.35-0.75), (0.1-0.35), respectively, detailed observations, are as follows:

- RSRP vs signal quality (RSRQ, SINR) correlation varies from moderate in urban area towards weaker in sub-urban area
- RSRP and RSRQ reveal higher correlation than RSRP with SINR, but because of the various factors that influence them differently, the relationship between RSRP and RSRQ is complex non-linear and changes from one deployment to the other. Their relationship is affected by environmental factors (e.g., urban vs. rural settings) and frequency bands used.
- RSRQ correlation with SINR reveal strongest dependency than any other signal pair, in any use case scenario
- RSRQ and SINR reveal comparable degree of impact on latency.

4.3.5 Data analysis summary

The approach undertaken for finding relationships among the collected data analysis led to the conclusion that RSRP has a distinct advantage in terms of metric stability and predictability. It tends to remain consistent (i.e. as depicted in Table 13, RSRP's median was observed similar in different areas), whereas RSRQ and SINR experience considerable fluctuations due to various environmental influences (e.g. SINR's median ranging from 6 to 12dB, with its values ranging from -10 to 35dB). In fact, the variations in SINR and RSRQ can be seen as overly dynamic and unsteady, even within the same deployment (the measured or reported values can differ based on the modem or carrier). Both RSRQ and SINR are important for assessing signal quality. They offer valuable insights into network quality and interference levels; however, their relevance may depend on the specific context and can sometimes be interchangeable. Conversely, RSRP emerges as a critical metric that must be prioritized in any optimization process, ensuring that the network delivers robust signal strength before fine-tuning based on SINR or RSRQ. Furthermore, RSRP vs signal quality (RSRQ, SINR) correlation varies and changes from one deployment to the other. The parameters relationship is complex and non-linear.

Based on the findings, it is assumed that for AI/ML algorithm applied for RAN Optimization must consider RSRP parameter, and the pair of parameters RSRQ and SINR cannot be used as alone input parameters for AI/ML model.

4.4 ML application

To validate ML applicability for the radio KPIs, the concept assumes RAN is the serving radio entity, where the 5G capable users are connected to, and deliver periodically the data determined based on the collected data sets, i.e.: RSRP, RSRQ, SINR. The system architecture, as illustrated in Figure 44, relies on the data sets as Input Data, that can be denoted as:

$$I = \{RSRP\ RSRQ, SINR\}, \tag{5}$$

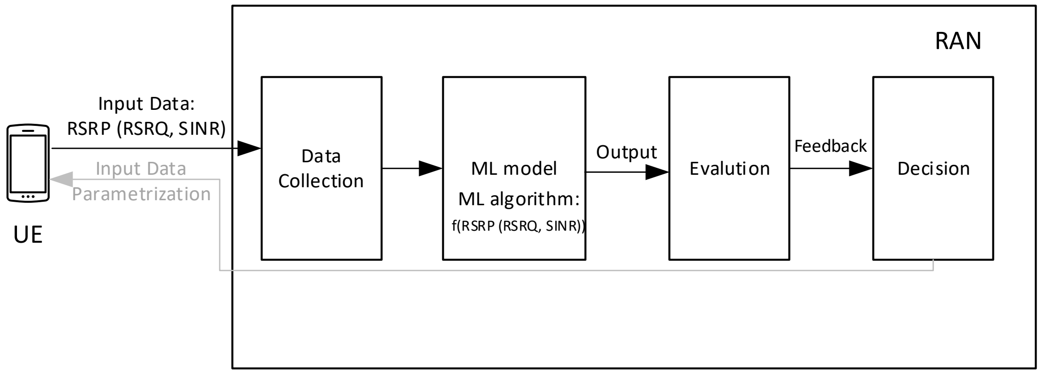


Figure 44: Scheme for applying AI/ML-based 5G RAN performance evaluation with RSRP, RSRQ, SINR parameters as Input Data.

Once made available to RAN (gNB), the Input Data emerges as Data Collection, that considers the signal strength metric: RSRP with highest importance, and signal quality input with consists of RSRQ and SINR metric, with lesser importance. The three parameters become target for the RAN to learn and predict the parameters, so that each metric collected as an individual entry is derived by an ML algorithm as a component of a predicted outcome dataset:

$$O = \{RSRP', RSRQ', SINR'\}. \tag{6}$$

The following sections validate and evaluate learning capabilities of a candidate ML algorithm, based on the findings learned from data collections analysis in live-network. The findings guide the algorithm selection as follows: RSRP, RSRQ, SINR are the primary units for data monitoring, but RSRP is the supreme characteristic and target for prediction. Any proven ML algorithm, with optimal performance, can be recommended to become integrated to RAN for enabling learning process, and thus paving the way for further optimizations.

For example, it is recognized that the Input Data parameters consist joint set of target KPIs for RAN Optimization use cases (Table 3), where Coverage, Mobility and QoS verification require these performance measurements to be taken for far-reaching optimization goals. Hence, beside the outcome evaluation of a ML algorithm, a follow-up objective is to identify potential implication and network operations enhancements arising from enabling ML-based monitoring of such dataset.

Namely, a potential overhead reduction technique for minimizing data amount is addressed in Chapter 5.4.

4.4.1 ML algorithm selection

In ML, a model refers to the mathematical representation of a data processing that exploits a set of rules and parameters that can be learned from the given data to make predictions. The process of creating a model involves training it on a dataset, where the model learns patterns and relationships within the data. Once trained, the model can be used to make predictions or classifications on new, unseen data. Out of the various types of ML models, that are broadly categorized into three main types: Supervised Learning, Unsupervised Learning and Reinforcement Learning Models.

For the identified Input Data Supervised Learning Model is assumed to be applicable type. This model requires a supervisor to determine input-output relationship to learn from labeled data. Due to known clustering of the radio KPIs to RSRP, RSRQ and SINR, and their normalized values, the validation exercise assume the model type suitability and supervision insurance by standard RAN operation (i.e. parameters values are expected with labels and within certain pre-defined range).

As theorized in [86], common supervised learning models imply mathematical algorithms using linear regression, decision trees, support vector machines, or neural networks.

4.4.1.1 Linear regression algorithm

Linear regression algorithms are based on the principle that the relationship between input features (I) and the target outputs (O) is a linear combination of the input feature (Figure 45: Principle for linear regression models.).The goal of the algorithm is to find the best linear relationship that can predict the output based in the input data. The algorithm tries to fit this straight line that best represents the relationship.

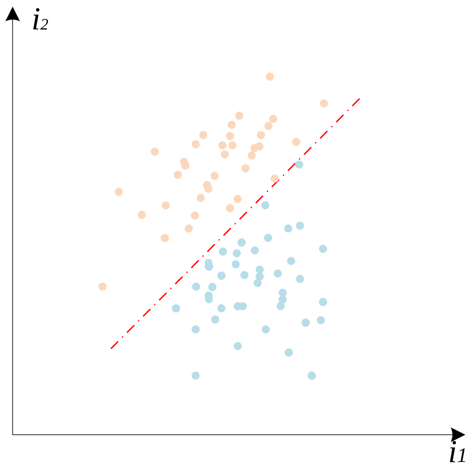


Figure 45: Principle for linear regression models.

Mathematically, the model can be expressed as:

$$o = w_1 i_1 + w_2 i_2 + .. + w_n i_n + b \quad (7)$$

where w represents weights and determine the influence of each feature on the output, the model “learns” the weights during training, and b is a bias parameter that allows model to adjust the output independently of the input features. The goal of training is to find the optimal weights w and bias b that minimize the difference between the model predictions and the actual outputs in the training data. The primary purpose of a linear or regression model is to predict continuous values.

4.4.1.2 Decision tree algorithm

A decision tree model in supervised learning is using an algorithm that builds a tree-like structure to make decisions or predictions based on input features. It works by recursively splitting the data into subsets, leading to a final prediction at the leaf nodes. Decisional trees are constructed by repeated splits of subsets of the original data into descendant subsets. The basic idea behind tree methods is that, based on the original data, a set of partitions is done so that the best class or value (in regression problems) can be determined. Despite being conceptually simple they are very powerful [86]. In a possible realization, the trees are binary (give responses that are nominal, such as 'true' or 'false'. Each step in a prediction involves checking the value of one predictor (variable). To predict, the algorithm starts at the top node, represented by a triangle (Δ) (Figure 46: Principle for decision tree models. The first decision compares two variables, depending on the outcome, it follows either the left or the right branch, and see that the tree classifies the data as type appropriately, and so on. To predict a response, the subsequent decisions are followed in the tree from the root (beginning) node down to a leaf node. The leaf node contains the response. Decision trees, depending on execution are called also classification trees and regression trees.

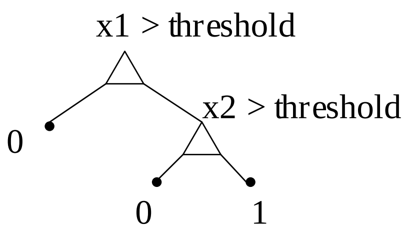


Figure 46: Principle for decision tree models.

4.4.1.3 Support Vector Machine algorithm

Support Vector Machine is supervised learning technique using a SVM classifier, that maps a set of inputs into a higher dimensional feature space. This is calculated through some linear or non-linear mapping. The algorithm objective is to maximize the distance between different classes and to find the hyperplane that produces the largest margin between different classes (Figure 47: Principle for SVM linear hyperplane [86]. [86].

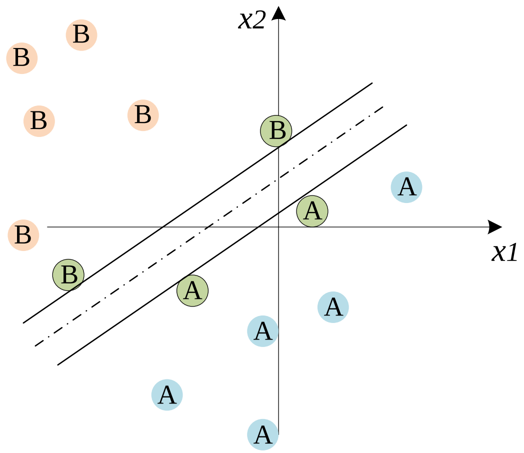


Figure 47: Principle for SVM linear hyperplane [86].

The SVM technique uses a subset of the training data as support vectors and they are crucial to the correct operation of this algorithm. In theoretical terms, the support vectors are the training samples that are closest to the decision surface and hence are the most difficult to classify. By finding the largest margin between these most difficult points, the algorithm can maximize the distance between classes and also guarantee that the decision region obtained for each class is the best one possible. For non-linear mapping, SVM can use different types of kernels, such as polynomial or Gaussian kernels [86].

For the collected data, the tree supervised algorithms have been applied to the collected data set, with the assumption that the algorithm suitable for non-linear dependencies, i.e. Decision Tree fits the data most.

4.4.2 ML model architecture

The architecture of the proposed ML model refers to an arrangement of its components, subsequent steps and the phases that enable the ML operation. The entire ML process for the experiment dataset consist of several steps, which can be split to:

- Input Data rendering: The Input Data I= {RSRP, RSRQ, SINR} needs to be formed as data sets for training and testing ML algorithm. For the proposed ML operation, it is assumed that training data contain three numerical variables (measurements results outcomes) recorded by the end user. Such structured data is given to an algorithm that performs observability process, during which the patterns and relationships from data are learnt and predictions are generated. Two approaches have been undertaken:
    - For the urban case scenario, dataset from MNO2 (where better signals propagation was observed) is split to partitions: Part of the samples are used as a training data to learn the model, whereas remaining portion of the dataset is used as new ‘unseen’ data.
    - For sub-urban scenario, dataset from MNO1 is split to partitions: Part of the samples are used as a training data to learn the model, whereas portion of the dataset is used as new ‘unseen’ data.
- Input Data Weighting: As an optional step, weights were applied to the input data, to reflect their importance. Two exercises were conducted: i) once RSRP variable was given a higher weight (0.7), indicating its greater importance compared to the other variables (RSRQ, SINR), ii) in another algorithm loop, SINR variable was given a higher weight (0.7), indicating its greater importance compared to the other variables.

- Training Models: Three supervised ML model types have been trained with MATLAB functions:
  - Linear Regression using fitlm function,
  - Decision Tree using fitctree function,
  - SVM using fitrsvm function.

During the training process, the models learned the optimal values for the input parameters. It is assumed that the learning process on given training data sets might support additional information and built-in features in MATLAB functions, such as customized weights and biases. The additional information are usually adjusted by supporting tools to minimize the difference between the model's predictions and the actual outcomes in the training data. With the available MATLAB configurations, the parameters that could be manually adjusted were weights associated with target parameter importance, whereas biases were left to toolbox implementation.

- Testing models: After training each of the models, the testing phase has been applied to 'unseen' datasets, to enable model performance check and adjustment. Here, it has been assumed that beyond given parameters as Input Data or learned parameters, models can have yet other parameters as hyperparameters, such as learning rate, accuracy, number of hidden layers, that can be implementation specific, but might have direct impact on training dynamics.
- Output Prediction: Trained and tested ML model processes the input features through learning, parameters patterns, and produces an output prediction. In the exercise, the generated Output Data was O = {RSRP', RSRQ', SINR'}. The nature of the output in this task is to predict numerical values.

4.4.3 ML model performance

The goal of ML model training is to find the optimal algorithm operational step (parameters selection, weights and bias) that minimize the difference between the model's predictions and the actual outputs in the training data. In supervised learning, the model is trained using a labeled dataset, which consists of input-output pairs, so that the input vector I= {RSRP, RSRQ, SINR}has a corresponding output vector O = {RSRP', RSRQ', SINR'}. To evaluate how well the trained and tested models perform and generalize the radio metrics monitoring and prediction task, The difference between model's prediction and actual outputs has been measured using common evaluation metric for regression models: R-Squared (R^2) which is a statistical measure that represents the proportion of the variance for a dependent variable that is explained by the independent variable(s) in the model. It provides an indication of how well the model fits the data. The mathematical formula for the measure can be represented as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi-\hat{y}i)^2}{\sum_{i=1}^{n}(yi-\bar{y})^2} \tag{8}$$

where:

- yi is the actual value for the i-th data point,
- $\hat{y}i$ is the predicted value for the i-th data point,
- $\bar{y}$ is the mean of the actual values yi,
- $\sum_{i=1}^{n}(yi-\hat{y}i)^2$ is the sum of squared errors (or residuals)
- $\sum_{i=1}^{n}(yi-\bar{y})^2$ is the total variation in the data, known as the Total Sum of Squares (TSS).

Range of R^2 values spans from 0 to 1, where the proportion interpretation of variability in the dependent variable that can be explained by the independent variables is defined as follows:

- R^2=1: The model perfectly fits the data (100% of the variance in the dependent variable is explained by the independent variable(s)).

- $R^2=0$: The model does not explain any of the variance (the model is no better than using the mean of the data).
- $R^2<0$: The model fits the data worse than a horizontal line at the mean of the data.

4.4.3.1 Results for Urban area

The training process outcome for the collected data set from urban scenario for the three applied ML models and calculated R^2 per model is illustrated in Figure 48, Figure 49 and Figure 50.

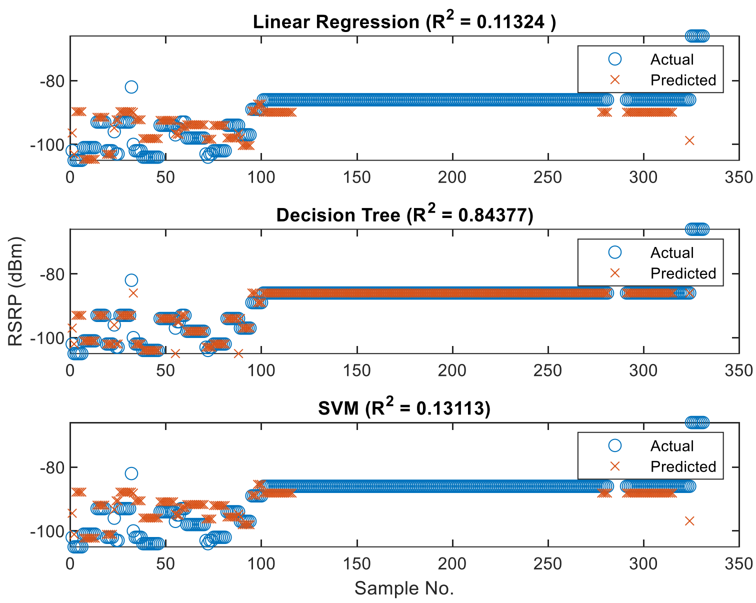


Figure 48: Linear Regression, Decision Tree and SVM models performance in urban scenario for RSRP

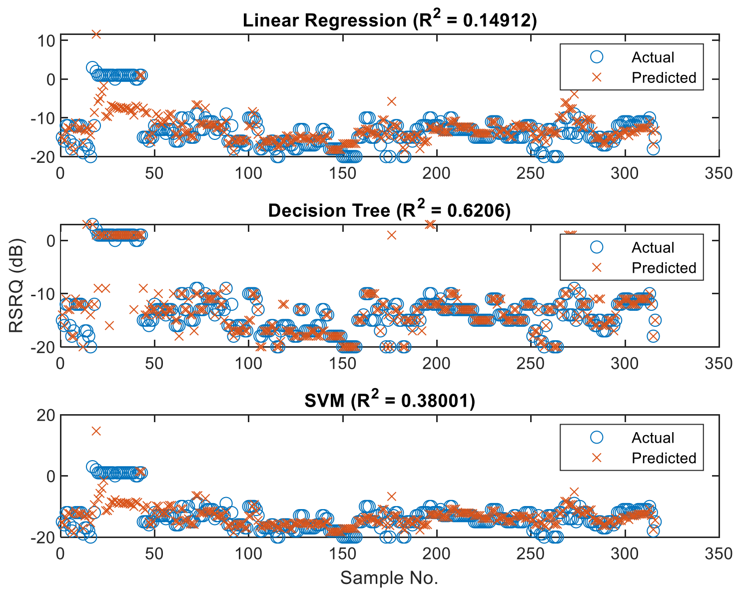


Figure 49: Linear Regression, Decision Tree and SVM models performance in urban scenario for RSRQ.

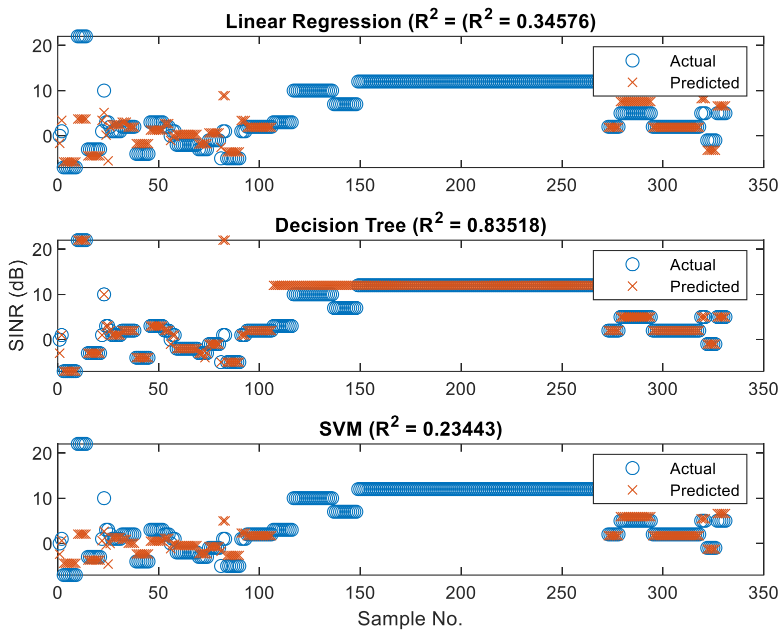


Figure 50: Linear Regression, Decision Tree and SVM models performance in urban scenario for SINR.

From 1.1k numerical samples for Input Data I={RSRP, RSRQ, SINR}, data has been split to training set ~700 samples and testing set ~330 samples. For Linear Regression and SVM models missing predicted values appeare, for each individual parameter, indicating a problem during the computation process and inability to perform prediction. Decision Tree model revealed capability to estimate the all the predictions and indicated R^2 value of 0.84, 0.62 and 0.83 for RSRP, RSRQ and SINR respectively. An R^2 value of 0.84, for RSRP and 0.83, for SINR suggest that the model explains 84% and 83% of the variance in the dependent variable, which can be considered as high performance and indicates that the model is capturing most of the underlying patterns in the data.

4.4.3.2 Results for Sub-urban area

The training process outcome for the collected data set from sub-urban scenario for the three applied ML models and calculated R^2 per model is illustrated in Figure 51, Figure 52 and Figure 53.

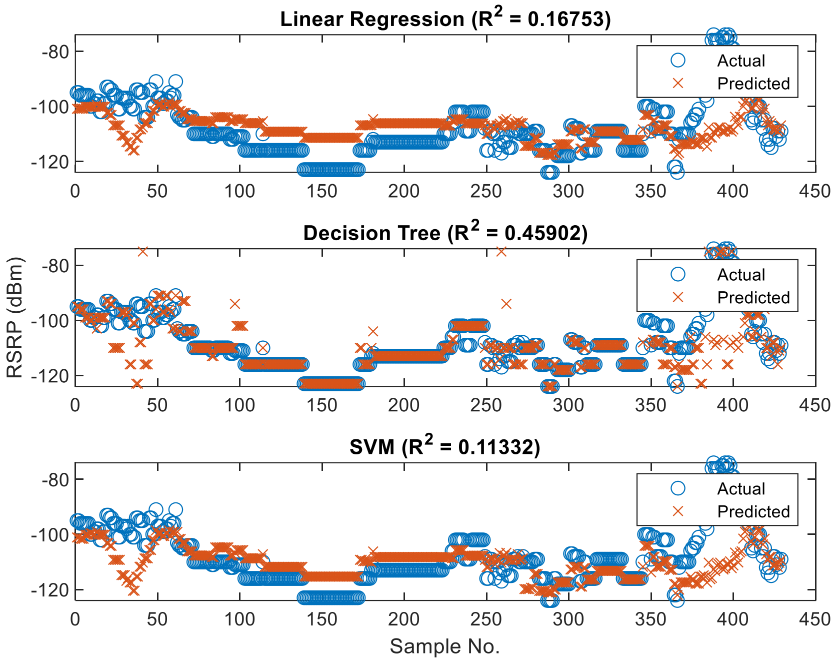


Figure 51: Linear Regression, Decision Tree and SVM models performance in sub-urban scenario for RSRP.

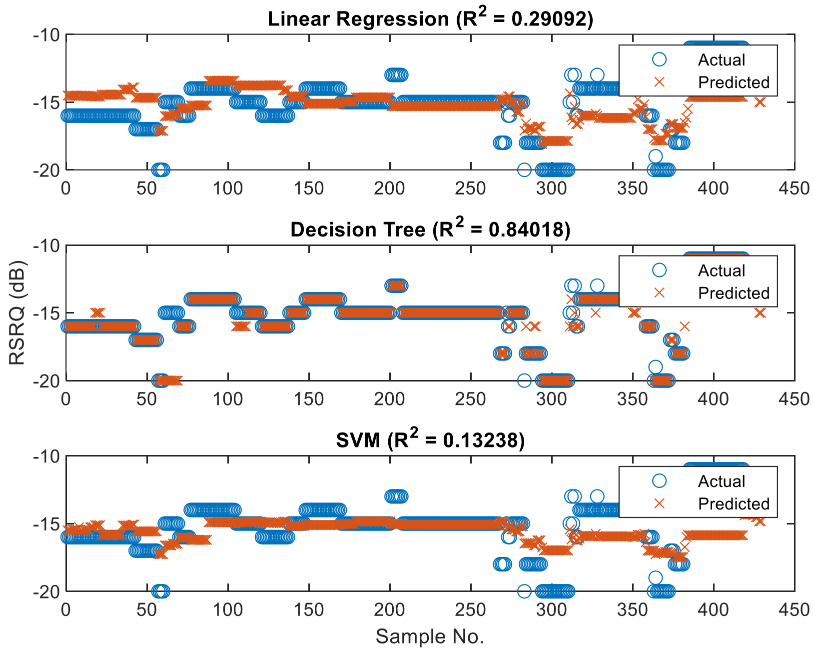


Figure 52: Linear Regression, Decision Tree and SVM models performance in sub-urban scenario for RSRQ.

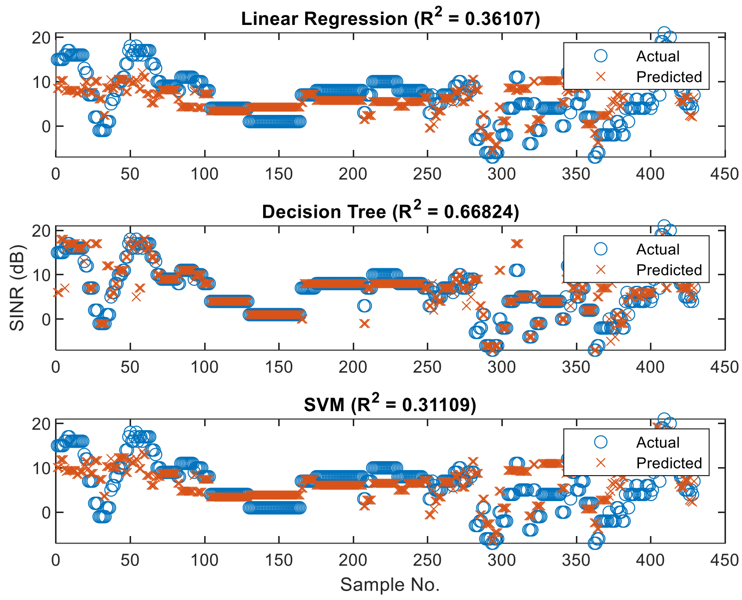


Figure 53: Linear Regression, Decision Tree and SVM models performance in sub-urban scenario for SINR.

From 1.2k numerical samples for Input Data I={RSRP, RSRQ, SINR}, data has been split to training set ~800 samples and testing set ~430 samples. All applied models performed learning and prediction, whereas Decision Tree model appeared to perform better for each individual parameter. R^2 value of 0.46, 0.84 and 0.66 for RSRP, RSRQ and SINR respectively, suggest that the model indcated higher performance for RSRQ and SINR, whereas encoutered difficulty in capturing majority of the underlying patterns in the data for RSRP.

To address the worsen performance, Input Data has been given to the algorithms with weighting the RSRP importance (assigning the weight 0.7), which resulted in a slight improvement of the Decision Tree model performance, as indicated in Figure 54.

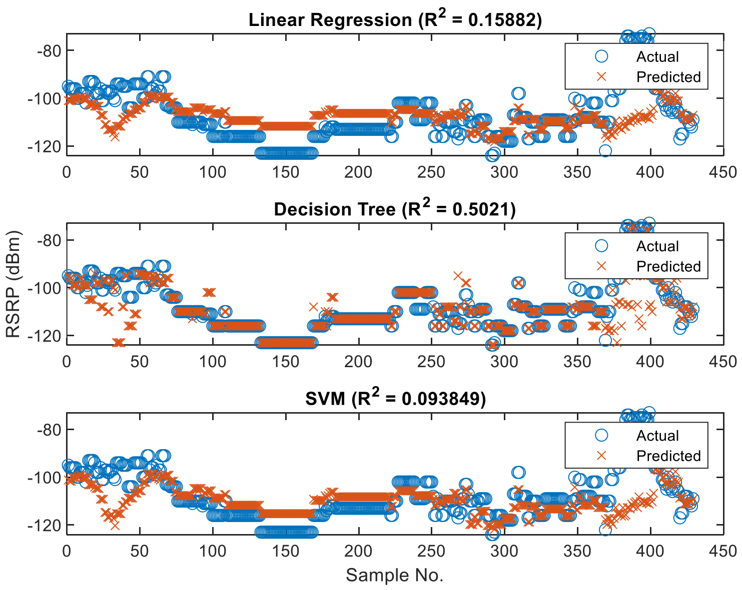


Figure 54: Linear Regression, Decision Tree and SVM models performance in sub-urban scenario for RSRP with weighted RSRP importance.

Decision Tree appears to be the model that is performing better than the other two models. However, compared to urban scenario, it seemed still underfitting and performing poorly on the tested data, in particular for RSRP. Hence, the amount of data used for training Decision Tree model has been increased by 20% samples, forcing the model to learn pattern closer to the data tail, which was tested with the model. This resulted in the performance increase to 63% for RSRP, as presented in Figure 55.

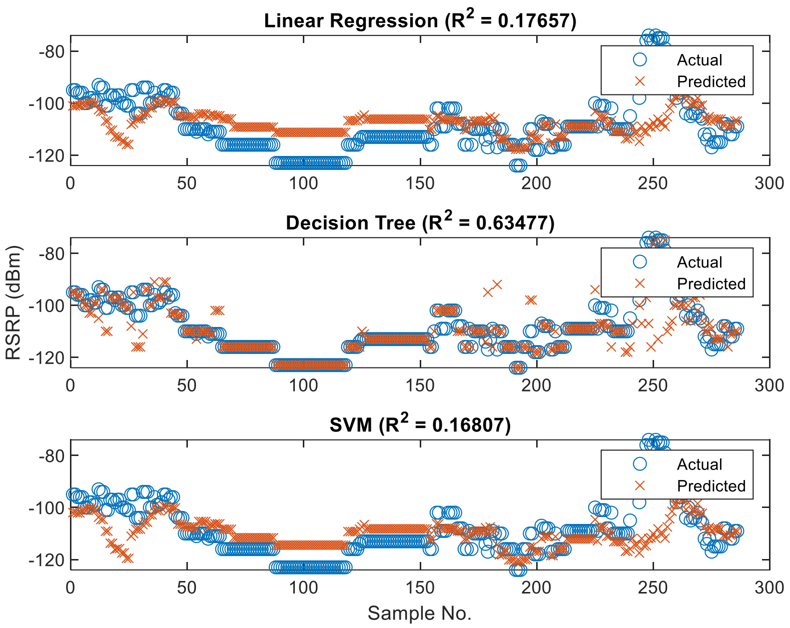


Figure 55: Linear Regression, Decision Tree and SVM models performance in sub-urban scenario for RSRP with increased training data set.

The result implies that the errors between the predicted and actual values became smaller, and thus, the model started to perform better.

4.4.4 Findings on ML application

This section tested the Linear Regression, Decision Tree, and SVM models applicability to the radio data set consisting of RSRP, RSRQ, SINR. The training process involved: fitting the model to the data, meaning the model selection and adjustments to minimize the error between its predictions and the actual outcomes. Essentially, RSRP, RSRQ, SINR became the input features to the models, and each of the models was trained and tested on how it learns parameters to generate a prediction for the input parameters. The learning process has been evaluated by comparing the model's predictions to the actual outcomes, using the evaluation metric R^2.
In general, ML model ability to fit depends on data dynamism, training scenario size, weights and other internal algorithm strategies. As expected here, for the key radio characteristics, which has no linear dependencies, and their dependencies vary, the Decision Tree ML model with least linear operation became most suitable. Linear Regression and SVM models weren't successful, whereas the performance of Decision Tree, with advanced regression, fit the examined data best.
The experiment identifies that linear approaches do not fit RAN observability. Decision Tree, as a model representing non-linear relationships, displays a higher R^2 and therefore confirms the model's effectiveness. Even equal weights applied to all three parameters (RSRP, RSRQ, SINR) have proved that the Decision Tree performance is best fitting out of the tested models. Although there are identified cases, where the model success is doubtful (Figure 51), a valuable approach for model

underfitting, can be weight adjustment or finding optimal training data size. The higher weight has been assigned to RSRP in the experiment, the better performance of the model could be observed.
Consequently, learning of the radio parameters has been proved with a conclusion that for the radio performance indicators, that are characterized by dependency on deployments or environmental conditions, iterative testing and monitoring for the variability impact is necessary. Once this is fundamental approach, fitting the model to the data can be more successful if the model is capable of handling more advanced operations than linear relationships among data. If a core algorithm of ML model operates based on regression with non-linear advancements (i.e., Decision tree), the model presents clear suitability, efficiency and scalability for radio metrics observation.

5 Strategy to implement

5.1 Foundations

Any 5G device needs to communicate with RAN to operate in the 5G network. When a UE connects to a gNB (next-generation Node B) in a 5G network, a series of messages are exchanged for successful communication, connection establishment and its maintenance Figure 56. Usually, after the initial RRC setup, the user provides feedback information to the gNB to keep the service at an adequate level.

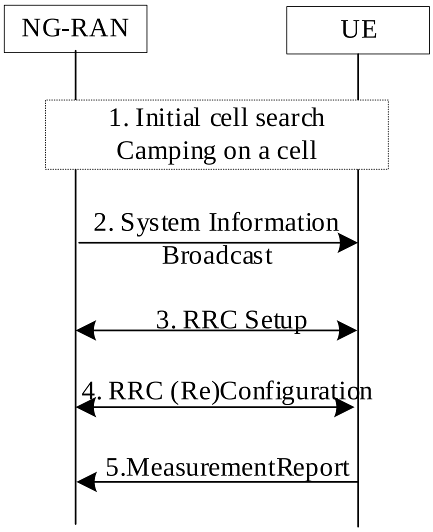


Figure 56: Typical RRC messages exchange between the UE and the gNB.

Step 4 configures a series of RRC protocol parameters, that determine the way on how the UE's feedback can be delivered. The key radio characteristics: RSRP, RSRP and SINR are measurement quantities, for which form of reporting and reporting periodicity is configured here. Once configured the UE measures the parameters and provides the report in Measurement Report accordingly. The mapping of the measurement reporting quantities to RRC reporting parameters value follows standardized resolution, which is specified for RSRP, RSRQ and SINR as presented in Table 14: RSRP measurement report quantity mapping [96].respectively.

Every UE, configured with RRC Reconfiguration message for periodical reporting of any measurement quantity out of RSRP, RSRQ, SINR, is triggered to pass the measurement result at time instances specified by the *ReportInterval* parameter. The parameter indicates the interval between periodical reports. The existing values for the periodicity ranges from 120ms to 60min [27]. This will result in the messages sequence as shown in Figure 57: RSRP, RSRQ, SINR reporting in air interface.

Table 14: RSRP measurement report quantity mapping [96].

Reported value	Measured quantity value (L3 SS-RSRP) and CSI-RSRP	Measured quantity value (L1 SS-RSRP and CSI-RSRP)	Unit
RSRP_0	SS-RSRP<-156	Not valid	dBm
RSRP_1	-156≤ SS-RSRP<-155	Not valid	dBm
RSRP_2	-155≤ SS-RSRP<-154	Not valid	dBm
RSRP_3	-154≤ SS-RSRP<-153	Not valid	dBm
RSRP_4	-153≤ SS-RSRP<-152	Not valid	dBm
RSRP_5	-152≤ SS-RSRP<-151	Not valid	dBm
RSRP_6	-151≤ SS-RSRP<-150	Not valid	dBm
…	…	…	
RSRP_16	-141≤ SS-RSRP<-140	RSRP<-140	dBm
RSRP_17	-140≤ SS-RSRP<-139	-140≤RSRP<-139	dBm
RSRP_18	-139≤ SS-RSRP<-138	-139≤ RSRP<-138	dBm
…	…	…	
RSRP_111	-46≤ SS-RSRP<-45	-46≤ RSRP<-45	dBm
RSRP_112	-45≤ SS-RSRP<-44	-45≤ RSRP<-44	dBm
RSRP_113	-44≤ SS-RSRP<-43	-44≤ RSRP	dBm
…	…		…
RSRP_127 (Note)	Infinity	Infinity	dBm
Note: The value of RSRP_127 is applicable for RSRP threshold configured by the network, but not for the purpose of measurement reporting.			

Table 15: RSRQ measurement report quantity mapping [96].

Reported value	Measured quantity value	Unit
RSRQ_0	RSRQ<-43	dB
RSRQ_1	-43≤ RSRQ<-42.5	dB
RSRQ_2	-42.5≤ RSRQ<-42	dB
RSRQ_3	-42≤ RSRQ<-41.5	dB
RSRQ_4	-41.5≤ RSRQ<-41	dB
..	..	…
RSRQ_122	17.5≤ RSRQ<18	dB
RSRQ_123	18≤ RSRQ<18.5	dB
RSRQ_124	18.5≤ RSRQ<19	dB
RSRQ_125	19≤ RSRQ<19.5	dB
RSRQ_126	19.5≤ RSRQ<20	dB
RSRQ_127	20 ≤ RSRQ	dB

Table 16: SINR measurement report quantity mapping [96].

Reported value	Measured quantity value (L3 SS-SINR and L3 CSI-SINR)	Measured quantity value (L1 SS-SINR and L1 CSI-SINR)	Unit
SINR_0	SINR<-23	SINR<-23	dB
SINR_1	-23≤ SINR<-22.5	-23≤SINR<-22.5	dB
SINR_2	-22.5≤ SINR<-22	-22.5≤SINR<-22	dB
SINR_3	-22≤ SINR<-21.5	-22≤SINR<-21.5	dB
SINR_4	-21.5≤ SINR<-21	-21.5≤SINR<-21	dB
..	..	..	…
SINR_123	38≤ SINR<38.5	38≤SINR<38.5	dB
SINR_124	38.5≤ SINR<39	38.5≤SINR<39	dB
SINR_125	39≤ SINR<39.5	39≤SINR<39.5	dB
SINR_126	39.5≤ SINR<40	39.5≤SINR<40	dB
SINR_127	40≤ SINR	40≤SINR	dB

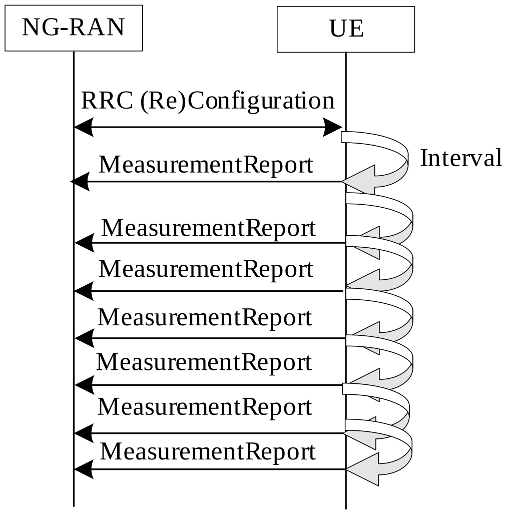


Figure 57: RSRP, RSRQ, SINR reporting in air interface.

As the reporting is in regular use in 5G network, the system supports enablers for the radio measurements automation through MDT (e.g. RRC configuration parameters to trigger the metrics collection). Automation of other type of data collection, within gNB (gNB metrics) is the key process of SON, which supports in addition placing some algorithms in RAN, however the algorithms themselves remain implementation specific. Based on the research and evaluations conducted in this dissertation, the approach on integrating the ML-based solutions to RAN monitoring integrates the two methods best practices: taking the automated data collection principled from MDT framework and inheriting placement of an algorithm in RAN. Reliance on ML algorithm, which operates on subsequent data records measuring radio performance does not limit the exploited methodology applicability to the network entity which receives the data (gNB). Likewise, it can be adopted to the transmitter side (UE), automating radio measurements generation and adding intelligence to radio metrics observability before transmitting the measurement results over the air. Although, it diversifies the ML technique adoption capabilities, it requires two possible approaches: the RAN-sided ML model and the UE-sided ML model with different operational and implementational impacts.

5.2 RAN-sided ML model

For the RAN-sided ML model, the assumption is to rely on automated data collection (MDT) through RRC signaling, SON capability to integrate optimization algorithms in NG-RAN, with minimum dedicated enhancements from the devices required. For that, a key implementation change is that a non-linear ML model fitting radio characteristic monitoring (e.g. with advanced regression methodology) is placed in the RAN. RAN collects series of RSRP, RSRQ, SINR, measurements reporting quantities, as per regular RRC operation, in a predefined report intervals. The collected entries are qualified as ML model Input Data. After activating, training and testing ML model, RAN is tasked to predict measurement quantities continuously, and evaluate them towards a decision on reporting parameters periodicity optimization. For instance, very satisfactory model performance with high R^2 value observed for a given time, might lead to a decision to release some of the metrics (e.g. one signal quantity measure, SINR or RSRQ, or both) and reconfigure the UE with coarser periodicity for some other measurement reporting (e.g. RSRP). This can be supported by re-use of the existing Report Interval with adjusted value to a longer interval. As an enhancement driven by a need to continuously monitor and refine the data collection process, it can be proposed that during monitoring phase, the UE provides the actual measurement quantity (e.g. for the quantities that have been turned off) to use it as a feedback for ML model and an input to re-evaluate the model performance (Figure 58: ML framework application to radio measurements monitoring with RAN-sided model.). That would ensure model adjustments and potential errors minimization. Without the UE signaling advancement, it is possible to apply the solution to all 5G devices. With the monitoring feedback, it is required to specify the supporting parameter in RRC signaling and rely on 5G devices implementing the extension.

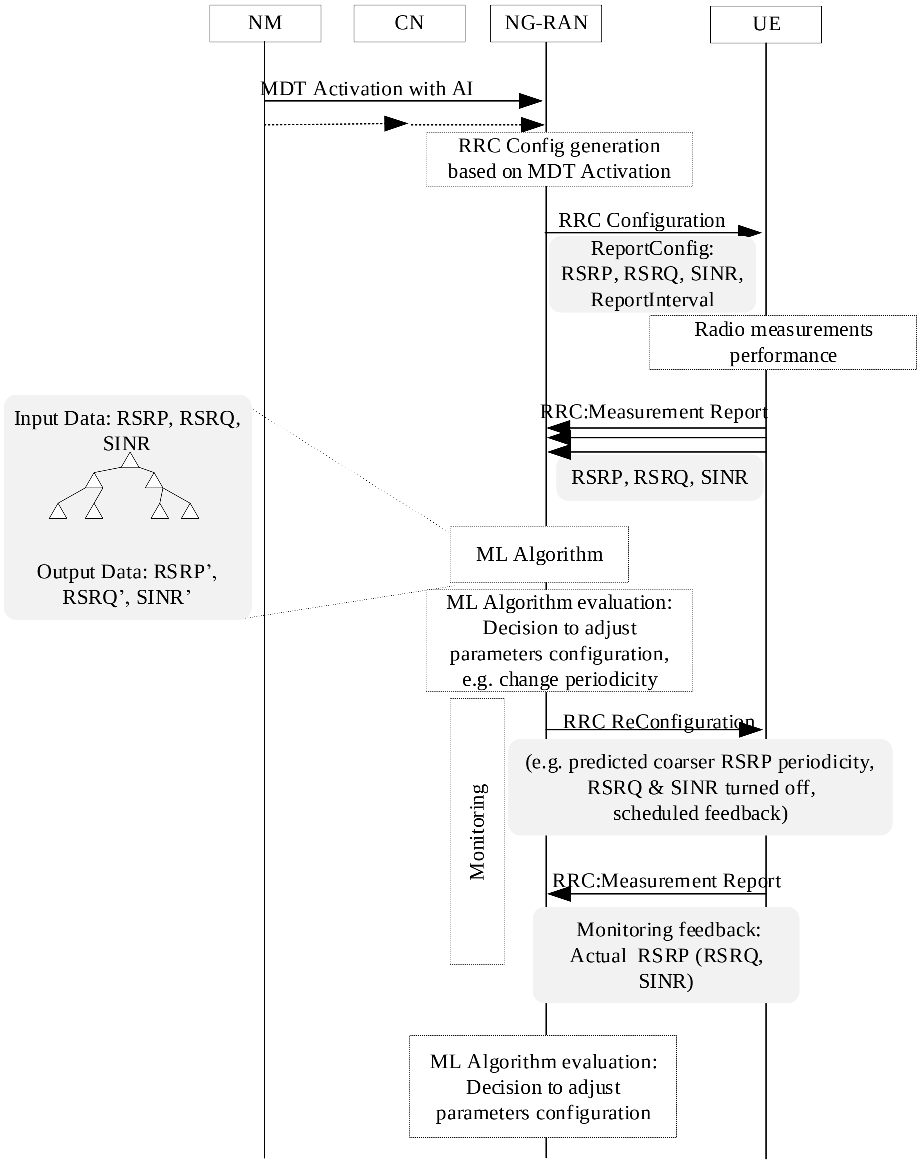


Figure 58: ML framework application to radio measurements monitoring with RAN-sided model.

5.3 UE-sided ML model

The other approach assumes that the UE implements ML model. Similarly, to the RAN-sided solution, the ML model capabilities fit radio measurements dataset variability. The key implementation dependence is that the ML model is placed in the UE. RAN needs to configure the series of RSRP, RSRQ, SINR, measurements reporting quantities, as per regular RRC operation, in a predefined report intervals. The measurement quantities are qualified internally within the device, as ML model Input Data. After activating, training and testing ML model, the UE is tasked to predict measurement quantities continuously, and evaluate them towards a decision on reporting parameters optimization. Simultaneously, to the ML operation the reporting of RSRP, RSRQ, SINR, measurements can proceed according to the configuration. However, the key value of applying ML model in the UE would be to enable the metrics reporting periodicity optimization, based on insights gained during ML model performance. To discourage negative impacts to the regular operations, from predicted reporting instances, it is assumed that the gNB should be aware and enabling start of the ML-assisted information exchange. This can be supported by simple extension (one bit flag) into RRC Reconfiguration message, that placed next to RSRP, RSRQ, SINR reporting configuration would indicate network readiness to coordinate data collection of measurement entities that result from ML model performance (e.g. ‘Enable Prediction’) (Figure **59**: ML framework application to radio measurements monitoring with UE-sided model.. Any well performing ML model for the generated measurement samples, could bring positive impact to radio interface by refining the measurements samples validation and providing evaluation outcomes that ultimately help to reduce signaling overhead (e.g. less measurement reporting quantities by coarser periodicity).

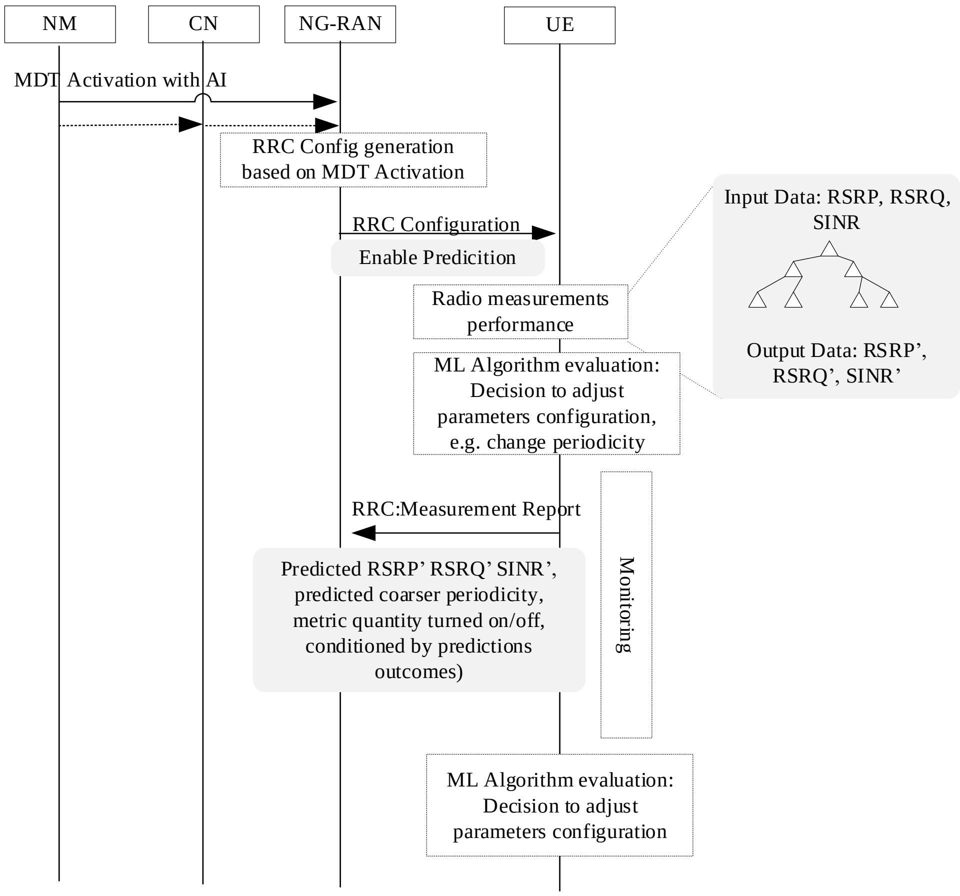


Figure 59: ML framework application to radio measurements monitoring with UE-sided model.

In this approach, it is critical to ensure attention and controllability over ML-enabled measurement reports, to avoid invalid or inaccurate entries reporting and to mitigate erroneous management of the radio resources. This RAN intelligence can be enabled by enhancement to air interface signaling (RRC), NG-RAN support and the UE enhanced implementation.

5.4 Benefits

In telecommunication network, overhead refers to data, resources and signaling required to maintain the connection. In radio interface, the overhead refers to the amount of data transfer, processing, and resources required to support the transmission between the UE and the gNB, over the radio. Regardless of the applied approach (NW-side or UE-side), augmenting the air-interface with ML-assisted radio measurements reporting offers significant radio interface overhead reduction, in particular referring to the amount of data transferred over the air. For understanding the level of gains, the data collection samples gathered during the experiment in this research work are taken as a reference, according to Table 17: Drive tests data collection samples relevant for mapping to air interface procedures..

Table 17: Drive tests data collection samples relevant for mapping to air interface procedures.

Data collection duration	Report periodicity	Measured quantities
30min	2 sec	RSRP, RSRQ, SINR

The RRC procedure that is used to transmit the measurement quantities entries of RSRP, RSRQ, SINR in UL is UL-DCCH: Measurement Report message. Total number of bits required over the air interface for the message transmission needs to consider: RRC message size, PDCP overhead, RLC overhead, MAC overhead and Physical layer overhead (CRC). For calculation simplicity, the overhead of lower layers is omitted, as it does not change nor contribute to the savings. The encoding of the Measurement Report message can vary from few to 120Bytes, depending on the information passed within it. In particular, the message that conveys only three measurement quantities: RSRP, RSRQ, SINR, using PER encoding [97], occupies 9Bytes (Figure 60: Measurement Report ASN.1 syntax encoding.

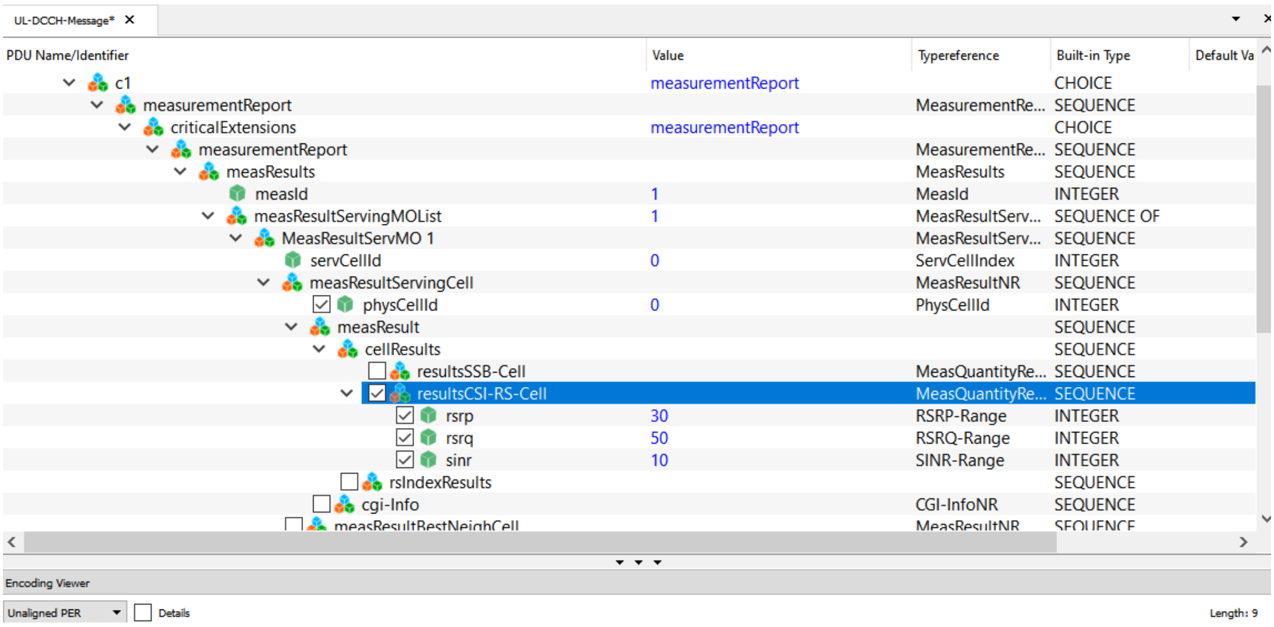


Figure 60: Measurement Report ASN.1 syntax encoding.

The data collection from a single user, with duration for 30min, and reporting RSRP, RSRQ, SINR every 4 seconds saves already 50% radio overhead, compared to reporting every 2 seconds, whereas changing the order of magnitude for report interval (e.g. from ms to sec, or from sec to min) can bring overhead reduction in the range from 98% to 75%.

Table 18: Drive tests data collection samples relevant for mapping to air interface procedures.

Data collection duration	Report periodicity	Measurement quantities	Reports amount	RRC messages size (Bytes)	RRC message size (bits)
n/a	n/a	RSRP, RSRQ, SINR	1	9	72
30min	640 ms	RSRP, RSRQ, SINR	2812	25308	202 463
30min	2 sec	RSRP, RSRQ, SINR	900	8100	64 800
30min	4 sec	RSRP, RSRQ, SINR	450	4050	32 400
30min	1 min	RSRP, RSRQ, SINR	30	270	2160

The considerations concern a single user. Scaling the reporting dimensions to a cell level (e.g. with hundreds of users) easily can visualize that ML-integrated into RAN, has great measurement reduction potential and can operate at a profit of radio overhead reduction. The measurement reduction potential can be statistically significant, however a validation of the overhead reduction gain needs to be proved in real-world environment, to ensure that the reduction is effective under various conditions.

To grasp the extent of the benefits, a simplified verification experiment has been conducted:

- for one dataset pair consisting of Input Data (i.e., the actually collected parameters by the UE (RSRP, RSRQ, SINR)) and the ML model outcome: Output Data (predicted RSRP, RSRQ, SINR), the number of samples amount has been cut to 50%, 75% and 98% of the original amount retrieved from a drive-test samples and data has been analyzed in context of accuracy.
- the reading of actually measured value has been compared to values predicted by ML algorithm with R-Squared (R^2) formula, used to measure the accuracy of the Decision Tree ML model's predictions for the reduced set of samples as shown in Table 19 and Figures 61, 62 and 63.

Table 19: ML Model validation for reduced number of samples.

Data collection duration		Number of samples trained by ML algorithm	Samples periodicity	Reduction in the number of samples	Real Reports amount	R^2
1	30min	900	2 sec	50%	450	0.67809
2	30min	450	4 sec	75%	225	0.67698
3	30min	30	1 min	98%	15	0.66969

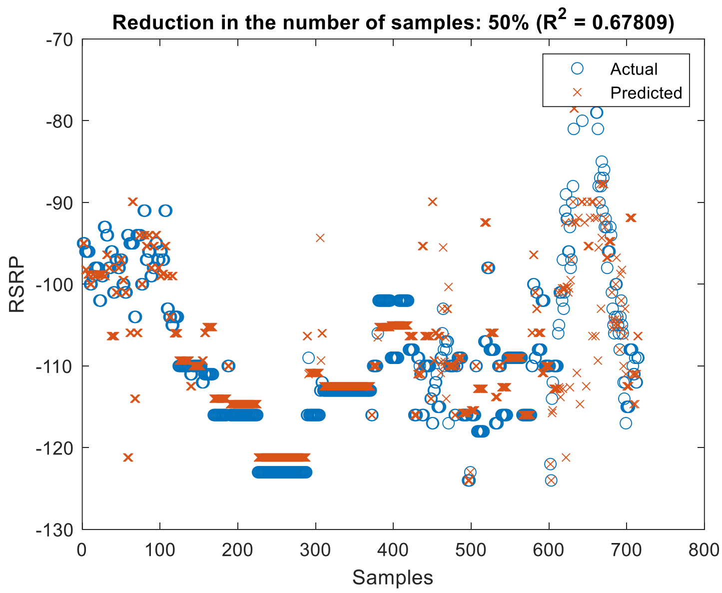


Figure 61: Decision Tree performance for number of samples reduced by 50%.

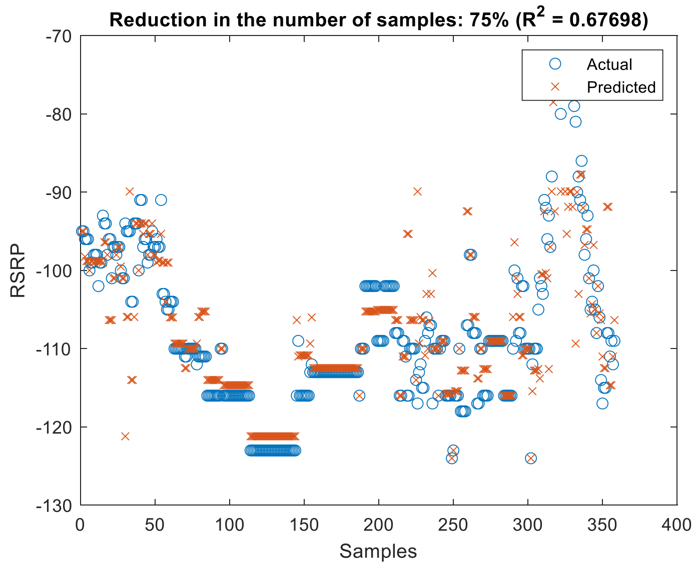


Figure 62: Decision Tree performance for number of samples reduced by 75%.

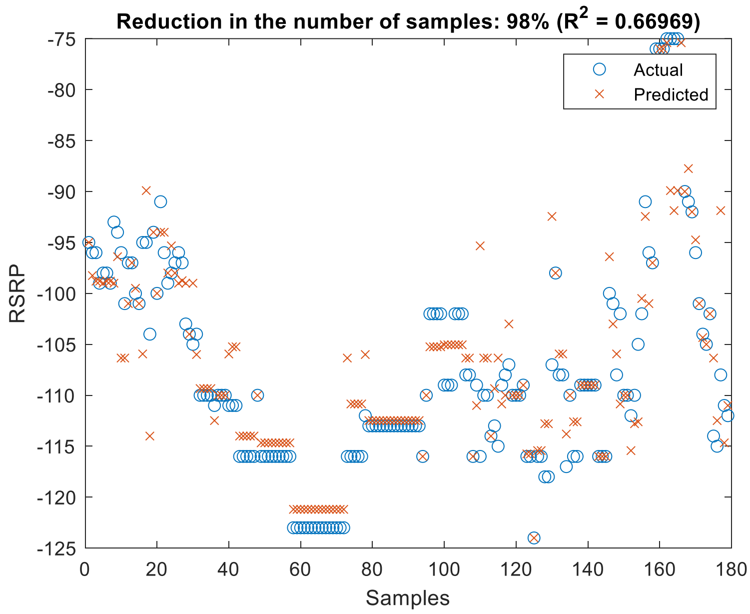


Figure 63: Decision Tree performance for number of samples reduced by 98%.

It can be observed that deviations appear regardless of a reduction rate. The R^2 starts to drop for higher overhead reductions indicating that predictions are closer to the actual values (meaning the model is performing well), if the samples reduction rate is lower.

Consequently, applying the overhead reduction technique should be jointly considered with accuracy and use case dynamizm (e.g. stationary UEs may achieve quantatively better reduction efficiency than mobile users, due to less changing environment). To ensure, that interplay between the potential overhead reduction gain and accuracy remains effective in real-world mobile network conditions or configurations changes, a continued monitoring over time is a fundamental requirement. If the overhead reduction technique is under regular validation to not compromise the user experience, the benefits derived from automated data collection through AI can enhance efficiency in network operations and initiate successful AI adoption in 5G.

6 Conclusion

In the advent of 5G networks, there is a recognized need for intelligent RAN operations and advanced automation. Telecommunications vendors that provide services and products to telecom operators (MNOs) are actively working on developing and implementing innovative solutions designed to support the rollout of 5G technology. These solutions incorporate a range of optimization strategies that have been utilized successfully in earlier generations of mobile networks (e.g. SON, MDT). They automate many human-involved intensive tasks for network operators, but once combined with higher dynamism and diversity, they become inadequate. The maintenance aspect of 5G presents a boosted level of complexity, resulting from diversified optimization targets and 5G requirements. Considering solely the three representative use cases within the framework of 5G technology: eMBB, URLLC, and mMTC, it becomes evident that the underlying supporting 5G infrastructure must provide a diverse range of functionalities. The system must meet often contradicting requirements (e.g. high data volume requirement for eMBB user may co-occur with ultra-low delay requirement for small data transmissions for URLLC devices). The recognition of these deficiencies, occurring simultaneously AI explosion, turns MNO industry, academia and telecommunication standards organizations attention to research on understanding AI applicability to 5G.

Guided by the trends and developments in the field, this dissertation attempted to determine actual feasibility, implications and gains from applying ML algorithm into RAN operations. In particular, by addressing the gap on applying ML algorithm to autonomous data collection and empowering it with self-learning and predictiveness capabilities. The goal, to address lack of methodology for identification of ML algorithm inputs, led to two key research questions:

i. which metrics in the massive amount of data are critical and can be considered as uniformly representative in context of the main co-existing 5G traffic use cases,
ii. which algorithm to select for the data processing in real-time operations with leveraged complexity, without imposing additional radio interface overhead?

Confronted with scientific literature and standardization developments directions, the research questions have evolved to two theses, namely:

1. Within radio characteristics set, RSRP is the metric with most important relevance for RAN monitoring.
2. Within ML algorithms set, Decision Tree proves to be suitable method for RAN monitoring in the context of overhead reduction.

Consequently, the thesis' strategy has been to first understand the practical scenarios in the the current 5G network rollouts and implications of the 5G network setup. The opening experiment has been to conduct manual drive tests campaigns in few deployment scenarios (urban, rural, highway area) for the 5G streamlined use case, i.e. eMBB. The empirical research aimed at uncovering how deployment and technological characteristics impact radio parameters importance based on a perception by the 5G-capable commercial devices. The statistical analysis on relationships among the collected data (RSRP, RSRQ, SINR) led to the conclusions that RSRP possess a clear advantage of the metric stability and predictiveness. It remains relatively stable, while RSRQ and SINR vary significantly due to many different environmental factors. In fact, SINR and RSRQ changes can be considered as too dynamic and unpredictable, even in the same deployment (measured or reported values vary by modem or carrier). Both: RSRP and SINR provide valuable insights into the signal quality and interference levels in the network, but their importance can be more context-dependent and, in some cases, interchangeable. Based on the findings offering validation of the Thesis 1, the research continued with the assumption that a ML algorithm applied for RAN monitoring must consider RSRP parameter, and the pair of parameters RSRQ and SINR cannot be used as alone input parameters for ML model.

When examining solely the streamlined 5G use case, represented by data collection from commercial 5G smartphones, namely eMBB, the conclusions can be considered with questioned applicability to URLLC, and mMTC. Though, the signal strength (RSRP) as the most basic and key radio performance metric is supposed to be measurable by any 5G device, by evident need that the elementary supporting 5G infrastructure must provide an adequate signal strength first. Therefore, relevance of the developed ML-enabled monitoring framework to the devices designed specifically for URLLC and mMTC might be considered as underlying, but guiding potentially additional validation and future research direction.

To validate Decision Tree-based ML applicability, the metrics combinations have been tested with Linear Regression, Decision Tree, and SVM algorithms with Matlab Toolbox. For this purpose, ML model undergone training process: fitting each model to the data and its adjustments to minimize the error between its predictions and the actual outcomes. Essentially, RSRP, RSRQ, SINR became the input features to the models, with RSRP weighed with higher importance. Each of the models was trained and tested on how it learns parameters to generate a prediction for the input parameters. The learning process has been evaluated by comparing the model's predictions to the actual outcomes, using the evaluation metric (R^2). The experiment identified that the more linear model, the less it fits RAN observability. Decision Tree, as a model representing non-linear relationships learning of the radio parameters has proved to be better performing and more successful, than linear models, validating Thesis 2.

Finally, the work has explored the strategy to implement the findings into live 5G network monitoring by extensive reuse of standardized 5G design, functioning signaling and interfaces, known from feasibility and implementation practices. The proposed framework integrates Decision Tree algorithm into 5G procedures relying on automated data collection from devices, and assuming the algorithm location in RAN, with two possible approaches: gNB-sided or UE-sided ML algorithm placement. As a result, the method enables signal strength metric (RSRP) collection with foremost importance for Decision Tree as ML algorithm use in RAN monitoring. It considers signal quality (RSRQ or SINR) as supporting metrics, that can be agilely observed for the variability impacts when necessary. Due to RSRP universal availability, applicability of the ML method is uniform, but not limited to further extensions, to any additional metric or metrics combinations by replacing or extension of the radio parameters datasets with other type of information. With such adaptive possibilities, great flexibility and scalability, the mechanism can serve a baseline to address a wide range of use cases, regardless of a traffic type and deployment scenario. In its presented form, the developed method facilitates integration of AI to 5G network with potential to reduce massive data collection production in RAN. Hence, it demonstrates practical viability of AI adoption into 5G RAN and beneficial trade-offs between possible complexity and ML-empowered automated optimization solutions in real-world mobile network.

Abbreviations

3GPP	3rd Generation Partnership
5GC	5g Core
AD	Anomaly Detection
AI	Artificial Intelligence
ANN	Artificial Neural Networks
AR	Augmented Reality
AS	Access Stratum
CF	Collaborative Filtering
CN	Core Network
CNN	Convolution Neural Networks
CP	Control Plane
CPICH	Common Pilot Channel Power
CPU	Central Processing Unit
CQI	Channel Quality Index
CSI	Chanel State Information
CU	Central Unit
D2D	Device-to-Device
DCCH	Dedicated Common Control Channel
DL	Deep Learning
DL	Downlink
DLR	Deep Reinforcement Learning
DPI	Deep Packet Inspection
DRX	Discontinuous Reception
DSP	Digital Signal Processing
DT	Decision Trees
DU	Distributed Unit
eMBB	Enhanced Mobile Broadband
EPC	Evolved Packet Core
FLOPS	Floating Point Operations Per Second
FR1	Frequency Range 1
FR2	Frequency Range 2
GA	Genetic Algorithms
gNB	5G NodeB
GNN	Graph Neural Network
GNN	Graph Neural Networks
GPU	Graphics Processing Unit
GSM	Global System for Mobile Communication
IMT	International Mobile Telecommunications
IoT	Internet of Things
ITU	International Telecommunication Union
KPI	Key Performance Indicator
KPI	Key Performance Indicator
LTE	Long Term Evolution
MAC	Medium Access Control
MCA	Multi-Dimensional Scaling
MCS	Modulation and Coding Scheme
MDT	Minimization of Drive Tests
MIMO	Multiple Input Multiple Output

ML	Machine Learning
mMTC	Massive Machine Type Communications
MNO	Mobile Networks Operators
MRO	Mobility Robustness
NF	Network Function
NG	Next Generation
NGMN	Next Generation Mobile Networks
NR	New Radio
NSA	Non-Stand Alone
OFDM	Orthogonal Frequency Division Multiplexing
OFDM	Orthogonal Frequency-Division Multiplexing
OFDMA	Orthogonal Frequency Domain Multiple Access
OPEX	Operational Expenditure
OPEX	Operational Expenditure
PCA	Principal Component Analysis
PCell	Primary Cell
PCI	Physical Cell Identity
PDCP	Packed Data Convergence Protocol
PER	Packet Encoding Rules
PHY	Physical
PRB	Physical Resource Block
QoE	Quality of Experience
QoS	Quality of Service
RACH	Random Access Channel
RAM	Random Access Memory
RAN	Radio Access Network
RB	Resource Block
RF	Radio Frequency
RL	Reinforcement Learning
RLC	Radio Link Control
RRC	Radio Resource Control
RRM	Radio Resource Management
RRM	Radio Resource Management
RS	Resynchronization Signals
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RSSI	Received Signal Strength Indication
SA	Stand Alone
SA	Standalone
SDAP	Service Data Adaptation Protocol
SINR	Signal to Interference plus Noise Ratio
SON	Self-Optimizing Networks
SS	Synchronization Signals
S-TMSI	S-Temporary Mobile Subscriber Identity
SVM	Support Vector Machine
SVM	Support Vector Machine
TCE	Trace Collection Entity
TS	Technical Specification
UE	User Equipment
UL	Uplink
UMTS	Universal Mobile Telecommunications System
UP	User Plane

URLLC	Ultra Reliable Low Latency Communications
VR	Virtual Reality
W	Watt
WCDMA	Wideband Code Division Multiple Access

List of tables

Table 1:Key performance indicators of 5G per traffic type. 18
Table 2: RAN Optimization Use Cases per method. 22
Table 3: Key performance indicators per RAN Optimization use case. 23
Table 4: Comparison of a supercomputer, personal computer; and the human brain processing capabilities, as per [32]. 24
Table 5: A comparison of traditional and ML approach, adopted from [33]. 26
Table 6: Comparison of Linear Regression, Decision Tree, and Support Vector Machine Methods. 29
Table 7: Detailed definition of key Radio performance indicators, as per [51-52]. 36
Table 8: Devices characteristics. 52
Table 9: Frequency bands detected in test case scenarios per operator. 52
Table 10: Exemplary device's desktop screens, with running 'Network Cell Info' Android application during drive test. 53
Table 11: Exemplary data logs. 57
Table 12: Radio performance indicators value ranges. 59
Table 13: Measurements results from the conducted test in different scenarios. 70
Table 14: RSRP measurement report quantity mapping [96]. 89
Table 15: RSRQ measurement report quantity mapping [96]. 89
Table 16: SINR measurement report quantity mapping [96]. 90
Table 17: Drive tests data collection samples relevant for mapping to air interface procedures. ... 94
Table 18: Drive tests data collection samples relevant for mapping to air interface procedures. ... 94
Table 19: ML Model validation for reduced number of samples. 95

List of figures

Figure 1: Overall 5G system architecture [19]. 11
Figure 2: NG-RAN architecture [26]. 12
Figure 3: 5G air interface. 13
Figure 4: User Plane protocol stack in radio [20]. 14
Figure 5: Control Plane protocol stack in radio [20]. 14
Figure 6: UE state machine and state transitions in NR [27]. 15
Figure 7: RRC connection establishment [27]. 16
Figure 8: RRC connection release [27]. 16
Figure 9: Usage scenarios for 5G and beyond [5]. 18
Figure 10: The importance of key capabilities in different 5G usage scenarios [5]. 19
Figure 11: SON architecture [29]. 20
Figure 12: SON functionalities and interactions within the network. 20
Figure 13: MDT architecture. 21
Figure 14: MDT framework. 22
Figure 15: Different machine learning approaches for data classification [41]. 27
Figure 16: Model of Artificial Neural Network. 28
Figure 17: MDT procedural steps. 37
Figure 18: Functional framework for RAN intelligence, source: [50]. 43
Figure 19. High level scheme for applying AI/ML-based 5G RAN performance evaluation based on 3GPP-assumed paradigm. 50
Figure 20: High level scheme for methodology undertaken in the research. 51
Figure 21: Signal range dependency on frequencies. 59
Figure 22: Distribution histogram of RSRP, RSRQ, SINR in urban area for MNO1. 60
Figure 23: Distribution of RSRP, RSRQ, SINR in urban area for MNO2. 61

Figure 24: Bar graph for distribution of RSRP in urban area for MNO1 (Nokia8.3) and MNO2 (NokiaG42). 62
Figure 25: Bar graph for distribution of SINR in urban area for MNO1 (Nokia8.3) and MNO2 (NokiaG42). 62
Figure 26: Bar graph distribution of SINR in urban area for MNO1 (Nokia8.3) and MNO2 (NokiaG42). 63
Figure 27: Distribution of RSRP, RSRQ, SINR in urban area for MNO3. 63
Figure 28: Bar graph for distribution of RSRP, RSRQ, SINR in urban area for MNO3. 64
Figure 29: Distribution of RSRP, RSRQ, SINR in sub-urban area for MNO1 (Nokia8.3). 65
Figure 30: Distribution of RSRP, RSRQ, SINR in sub-urban area for MNO1 (NokiaG42). 65
Figure 31: Bar graph for distribution of RSRP in sub-urban area for MNO1 (Nokia8.3, NokiaG42). 66
Figure 32: Bar graph for distribution of RSRQ in sub-urban area for MNO1 (Nokia8.3, NokiaG42). 66
Figure 33: Bar graph for distribution of SINR in sub-urban area for MNO1 (Nokia8.3, NokiaG42). 67
Figure 34: Loss of 5G signal in sub-urban area by Nokia8.3. 68
Figure 35: Distribution of RSRP, RSRQ, SINR in a highway area for MNO1. 69
Figure 36: Bar graph for distribution of RSRP, RSRQ, SINR in a highway area for MNO1. 69
Figure 37: Block model for computing cross-correlation. 72
Figure 38: Correlation co-efficient for RSRP, RSRQ, SINR in Urban area, MNO1 (Nokia8.3). 73
Figure 39: Correlation co-efficient for RSRP, RSRQ, SINR in Urban area, MNO2 (NokiaG42). 73
Figure 40: Correlation co-efficient for RSRP, RSRQ, SINR in sub-urban area, MNO1 (Nokia8.3). 74
Figure 41: Correlation co-efficient for RSRP, RSRQ, SINR in sub-urban area, MNO1 (NokiaG42). 74
Figure 42: Correlation co-efficient for user's throughput with RSRP, RSRQ and SINR in urban area, MNO3. 75
Figure 43: Correlation co-efficient for user's latency with RSRP, RSRQ and SINR in urban area, MNO3. 76
Figure 44: Scheme for applying AI/ML-based 5G RAN performance evaluation with RSRP, RSRQ, SINR parameters as Input Data. 77
Figure 45: Principle for linear regression models. 78
Figure 46: Principle for decision tree models. 79
Figure 47: Principle for SVM linear hyperplane [86]. 80
Figure 48: Linear Regression, Decision Tree and SVM models performance in urban scenario for RSRP. 82
Figure 49: Linear Regression, Decision Tree and SVM models performance in urban scenario for RSRQ. 82
Figure 50: Linear Regression, Decision Tree and SVM models performance in urban scenario for SINR. 83
Figure 51: Linear Regression, Decision Tree and SVM models performance in sub-urban scenario for RSRP. 84
Figure 52: Linear Regression, Decision Tree and SVM models performance in sub-urban scenario for RSRQ. 84
Figure 53: Linear Regression, Decision Tree and SVM models performance in sub-urban scenario for SINR. 85
Figure 54: Linear Regression, Decision Tree and SVM models performance in sub-urban scenario for RSRP with weighted RSRP importance. 85
Figure 55: Linear Regression, Decision Tree and SVM models performance in sub-urban scenario for RSRP with increased training data set. 86
Figure 56: Typical RRC messages exchange between the UE and the gNB. 88

Figure 57: RSRP, RSRQ, SINR reporting in air interface. 90
Figure 58: ML framework application to radio measurements monitoring with RAN-sided model. 92
Figure 59: ML framework application to radio measurements monitoring with UE-sided model.. 93
Figure 60: Measurement Report ASN.1 syntax encoding. 94
Figure 61: Decision Tree performance for number of samples reduced by 50%. 95
Figure 62: Decision Tree performance for number of samples reduced by 75%. 96
Figure 63: Decision Tree performance for number of samples reduced by 98%. 96

References

1. Rochman, M. I., Sathya, V., Fernandez, D., Nunez, N., Ibrahim, A. S., Payne, W., & Ghosh, M. (2023). A comprehensive analysis of the coverage and performance of 4G and 5G deployments. Computer Networks, 237. https://doi.org/10.1016/j.comnet.2023.110060
2. 5G core deployments worldwide from 2018 to 2023, by type, Statista, https://www.statista.com/statistics/1330511/5g-core-deployments-worldwide-by-type/
3. ITU, Measuring digital development, Fact and figures 2023, https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx
4. ITU, 5G - Fifth generation of mobile technologies, https://www.itu.int/en/mediacentre/backgrounders/Pages/5G-fifth-generation-of-mobile-technologies.aspx
5. ITU-R M.2083-0: IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond
6. ITU-R M.1645: Framework and overall objectives of the future development of IMT-2000 and system beyond IMT-2000
7. ITU-R M.2150-1: Detailed specifications of the terrestrial radio interfaces of International Mobile Telecommmunications-2020 (IMT-2020)
8. 3GPP TS 36.902 Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Self-Configuring and Self-Optimizing Network (SON) Use Cases and Solutions, https://www.3gpp.org/ftp/Specs/archive/36_series/36.902/36902-931.zip
9. 3GPP TS37.320, Radio measurement collection for Minimization of Drive Tests (MDT); Overall description; Stage 2, https://www.3gpp.org/ftp/Specs/archive/37_series/37.320/37320-i20.zip
10. Qu, M., Tang, X., Zhang, Y., Chen, Z., Zhang, T., Zhao, Y., & Li, Y. (n.d.). An Overview of Enabling Artificial Intelligence in 3GPP 5G-Advanced Networks. https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics60724.2023.00035
11. Michelle Lam, Intelligent automation strategies for next-generation RAN, Analysys Mason, 2024
12. Chen H., Success factors impacting artificial intelligence adoption: Perspective form the Telecom Industry in China, 2019
13. AI Revolution in Telecommunications: Transforming Connectivity, https://www.vertis.com, 2024
14. 3GPP, Finding AI in 3GPP, https://www.3gpp.org/technologies/finding-ai-in-3gpp, 2022
15. RP-221248, 3GPP Study on AI-ML for air interface, http://3gpp.org/ftp/tsg_ran/TSG_RAN/TSGR_96/Docs/RP-221248.zip
16. RP-170855, "New WID on New Radio Access Technology," 3GPP RAN #75, Dubrovnik, Croatia, March 2017; http:// www.3gpp.org/ftp/TSG_RAN/TSG_RAN/TSGR_75/Docs/ RP-170855.zip

17. 3GPP TR 38.913, Study on Scenarios and Requirements for Next Generation Access Technologies, V18.0.0, April 2024, https://www.3gpp.org/ftp/Specs/archive/38_series/38.913/38913-i00.zip
18. 3GPP TR 38.912, Study on new radio access technology,V18.0.0, April 2024, https://www.3gpp.org/ftp/Specs/archive/38_series/38.912/38912-i00.zip
19. 3GPP TS23.501, System architecture for the 5G System, https://www.3gpp.org/ftp/Specs/archive/23_series/23.501/23501-j00.zip
20. 3GPP TS38.300, NR; NR and NG-RAN Overall description; Stage-2, https://www.3gpp.org/ftp/Specs/archive/38_series/38.300/38300-i20.zip
21. ITU-R M.2410: Minimum requirements related to technical performance for IMT-2020 radio interface(s)
22. ITU-R M.2412: Guidelines for evaluation of radio interface technologies for IMT-2020
23. Akyildiz, I. F., Nie, S., Lin, S. C., & Chandrasekaran, M. (2016). 5G roadmap: 10 key enabling technologies. Computer Networks, 106, 17–48.
24. Takeda, Kazuki & Xu, Huilin & Kim, Taehyoung & Schober, Karol & Lin, Xingqin. (2019). Understanding the Heart of the 5G Air Interface: An Overview of Physical Downlink Control Channel for 5G New Radio (NR), https://www.researchgate.net/publication/336303851
25. Lin, X., Grovlen, A., Werner, K., Li, J., Baldemair, R., Cheng, J. F. T., Parkvall, S., Larsson, D. C., Koorapaty, H., Frenne, M., & Falahati, S. (2019). 5G New Radio: Unveiling the Essentials of the Next Generation Wireless Access Technology. IEEE Communications Standards Magazine, 3(3), 30–37. https://doi.org/10.1109/MCOMSTD.001.1800036
26. 3GPP TS38.401, NG-RAN; Architecture description, https://www.3gpp.org/ftp/Specs/archive/38_series/38.401/38401-i20.zip
27. 3GPP TS38.331, NR; Radio Resource Control (RRC); Protocol Specification, https://www.3gpp.org/ftp/Specs/archive/38_series/38.331/38331-i20.zip
28. NGMN, NGMN Recommendation on SON and O&M Requirements
29. 3GPP TS28.313, Management and orchestration; Self-Organizing Networks (SON) for 5G networks, https://www.3gpp.org/ftp/Specs/archive/28_series/28.313/28313-j00.zip
30. 3GPP TS32.422, Telecommunication management; Subscriber and equipment trace; Trace control and configuration management, https://www.3gpp.org/ftp/Specs/archive/32_series/32.422/32422-i30.zip
31. Jurafsky, Dan, Richardson, Daniel. "Symbolic Systems 100: Introduction to Cognitive Science", 2005
32. Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach, 4th edition
33. Artificial Intelligence and Big Data: The Birth of a New Intelligence, First Edition. Fernando Iafrate, 2018. Published by ISTE Ltd and John Wiley & Sons, Inc.
34. Searle, John R (1980), Minds, brains, and programs, Behavioral and Brain Sciences, https://courses.washington.edu/info300/searle.pdf
35. Flowers, J. C. "Strong and Weak AI: Deweyan Considerations." Worcester State University, Worcester, Massachusetts. Available at: ceur-ws.org/Vol-2287/paper34.pdf
36. Stanford Encyclopedia of Philosophy, The Chinese Room Argument, https://plato.stanford.edu/entries/chinese-room/
37. Boucher, Philip. "Artificial intelligence: How does it work, why does it matter, and what can we do about it?", June 2020, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU(2020)641547_EN.pdf
38. Liu, B., "Weak AI is Likely to Never Become Strong AI, So What is its Greatest Value for us?", 2021, arxiv.org/pdf/2103.15294
39. Cardona, Miguel A., Rodríguez, Roberto J., Ishmael, Kristina. "Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations". 2023. www2.ed.gov/documents/ai-report/ai-report.pdf

40. Mohammed M, Khan MB, Bashier Mohammed BE, Machine learning: algorithms and applications. CRC Press; 2016
41. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160, 2021, https://doi.org/10.1007/s42979-021-00592-x
42. George A.F. Seber, Alan J, Lee, Linear Regression Analysis, Wiley
43. RP-221348, Study on AI-ML for NR air interface, 3GPP RAN#96, Budapest, June 2022
44. http://3gpp.org/ftp/tsg_ran/TSG_RAN/TSGR_96/Info_for_workplan/revised_SIDs_11/RAN1_7/RP-221348.zip
45. 3GPP TS51.010-1, Mobile Station (MS) confirmance specification; Part 1: Conformance Specification, https://www.3gpp.org/ftp/Specs/archive/51_series/51.010-1/51010-1-de0.zip
46. Harri Holma, Antti Toskala, WCDMA for UMTS: HSPA Evolution and LTE, John Wiley & Sons, Ltd, doi: 10.1002/9780470669501
47. R2-093965, Justification of measurements for drive tests, 3GPP RAN2#66bis, NTT Docomo Inc.
48. Harri Holma, Antti Toskala, LTE for UMTS, Evolution to LTE-Advanced, John Wiley & Sons, Ltd, doi: 10.1002/9780470669501
49. 3GPP TS32.425, Telecommunication management; Performance Management (PM); Performance measurements, E-UTRAN, https://www.3gpp.org/ftp/Specs/archive/32_series/32.425/32425-i00.zip
50. 3GPP TR 37.817, V1.0.0 Study on enhancement for data collection for NR and ENDC
51. 3GPP TS 36.214, E-UTRA, Physical layer, Measurements, https://www.3gpp.org/ftp/Specs/archive/36_series/36.214/36214-i00.zip
52. 3GPP TR 38.215, NR, Physical layer, Measurements, https://www.3gpp.org/ftp/Specs/archive/38_series/38.215/38215-i30.zip
53. 3GPP TS 36.211, E-UTRA, Physical channels and modulation, https://www.3gpp.org/ftp/Specs/archive/36_series/36.211/36211-i01.zip
54. 3GPP TS 38.211, NR; Physical channels and modulation, https://www.3gpp.org/ftp/Specs/archive/38_series/38.211/38211-i30.zip
55. Afroz, F., Subramanian, R., Heidary, R., Sandrasegaran, K., & Ahmed, S. (2015). SINR, RSRP, RSSI and RSRQ Measurements in Long Term Evolution Networks. International Journal of Wireless & Mobile Networks, 7(4), 113–123. https://doi.org/10.5121/ijwmn.2015.7409
56. Jay Weitzen, Rachel Wakim, Erin Webster, Comparing RSRP, CQI, and SINR measurements with Predictions for Coordinated and Uncoordinated LTE small cell Networks, Microwaves, Communications, Antennas and Electronic Systems (COMCAS), 2015 IEEE International Conference on : date 2-4 Nov. 2015. (2015). IEEE.
57. Weitzen, Jay A. and Grosch, Theodore, "Comparing Coverage Quality for Femtocell and Macrocell Broadband Data Services", IEEE Communications Magazine, January 2010
58. AWE Communications, Germany. Software tool WinProp for the planning of mobile communication networks,; 2013
59. J. Kurjenniemi and T. Henttonen, "Effect of measurement bandwidth to the accuracy of inter-frequency RSRP measurements in LTE," 2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications, Cannes, France, 2008, pp. 1-5, doi: 10.1109/PIMRC.2008.4699694.
60. J. Puttonen, J. Turkka, O. Alanen and J. Kurjenniemi, "Coverage optimization for Minimization of Drive Tests in LTE with extended RLF reporting," 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Istanbul, Turkey, 2010, pp. 1764-1768, doi: 10.1109/PIMRC.2010.5671907.
61. Oliosi, E., Mathew, S., Davoli, L., Ferrari, G., Strozzi, N. O., Manghi, F., Casoli, F., & Notari, A. (n.d.). Simulation-based Analysis of Experimental 5G NR Downlink CSI-RSRP-based Handover Performance.

62. Rochman, M. I., Sathya, V., Fernandez, D., Nunez, N., Ibrahim, A. S., Payne, W., & Ghosh, M. (2023). A comprehensive analysis of the coverage and performance of 4G and 5G deployments. Computer Networks, 237. https://doi.org/10.1016/j.comnet.2023.110060
63. Choi, Y. H. (2024). Adaptive beamforming for maximum SINR in the presence of correlated interferences. Digital Signal Processing: A Review Journal, 146. https://doi.org/10.1016/j.dsp.2023.104370
64. Patel, W., & Tripathy, P. (2023). eMBB, URLLC, and mMTC 5G Wireless Network Slicing: A Communication-Theoretic View. 2023 International Conference on Artificial Intelligence and Smart Communication, AISC 2023, 1291–1296. https://doi.org/10.1109/AISC56616.2023.10085623
65. Lin, X., Grovlen, A., Werner, K., Li, J., Baldemair, R., Cheng, J. F. T., Parkvall, S., Larsson, D. C., Koorapaty, H., Frenne, M., & Falahati, S. (2019). 5G New Radio: Unveiling the Essentials of the Next Generation Wireless Access Technology. IEEE Communications Standards Magazine, 3(3), 30–37. https://doi.org/10.1109/MCOMSTD.001.1800036
66. A. A. Esswie and K. I. Pedersen, "Opportunistic Spatial Preemptive Scheduling for URLLC and eMBB Coexistence in Multi-User 5G Networks," in IEEE Access, vol. 6, pp. 38451-38463, 2018, doi: 10.1109/ACCESS.2018.2854292.
67. 3GPP Release 15, https://www.3gpp.org/specifications-technologies/releases/release-15
68. 3GPP TS22.261, Service requirements for the 5G system, https://www.3gpp.org/ftp/Specs/archive/22_series/22.261/22261-j70.zip
69. SP-220439, Study on AI/ML Model Transfer Phase 2, https://www.3gpp.org/ftp/tsg_sa/TSG_SA/TSGS_96_Budapest_2022_06/Docs/SP-220439.zip
70. 3GPP TS28.104, Management and orchestration; Management Data Analytics (MDA)
71. 3GPP TS28.105, Management and orchestration; Artificial Intelligence/Machine Learning (AI/ML) management
72. 3GPP TR 28.908, Study on Artificial Intelligence/Machine Learning (AI/ML) management
73. 3GPP, AI/ML management, https://www.3gpp.org/technologies/ai-ml-management
74. 3GPP TR23.700-80, Study on 5G System Support for AI/ML-based Services
75. RP-240323, Study on enhancements for Artificial Intelligence (AI)/Machine Learning (ML) for NG-RAN
76. RP-234039, Artificial Intelligence(AI)/Machine Learning (ML) for NR Air Interface, RAN#102, December, 2023
77. TR38.843, Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR air interface.
78. Qu, M., Tang, X., Zhang, Y., Chen, Z., Zhang, T., Zhao, Y., & Li, Y. (n.d.). *An Overview of Enabling Artificial Intelligence in 3GPP 5G-Advanced Networks*. https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics60724.2023.00035
79. W. Chen *et al.*, "AI assisted PHY in future wireless systems: Recent developments and challenges," in *China Communications*, vol. 18, no. 5, pp. 285-297, May 2021, doi: 10.23919/JCC.2021.05.019.
80. He S, Xiong S, Ou Y, et al. An overview on the application of graph neural networks in wireless networks. IEEE Open Journal of the Communications Society, 2021.
81. Mohammed M, Khan MB, Bashier Mohammed BE, Machine learning: algorithms and applications. CRC Press; 2016
82. Cheung, B., Kumar, G., & Rao, S. A. (2005). Statistical algorithms in fault detection and prediction: Toward a healthier network. *Bell System Technical Journal*, *9*(4), 171–185. https://doi.org/10.1002/bltj.20070
83. Thamilchelvan, R., & Gomathy, C. (2023). Modeling Intelligent Wireless Communication Channel for 5G and Beyond using Advanced Machine Learning Techniques. *Proceedings of the 5th International Conference on Inventive Research in Computing Applications, ICIRCA 2023*, 916–920. https://doi.org/10.1109/ICIRCA57980.2023.10220931

84. Zhang, J., Zheng, G., Zhang, Y., Krikidis, I., & Wong, K.-K. (2023). *Deep Learning Based Predictive Beamforming Design*. https://doi.org/10.1109/TVT.2023.3238108
85. Dai, L., Jiao, R., Adachi, F., Poor, H. V., & Hanzo, L. (2020). Deep Learning for Wireless Communications: An Emerging Interdisciplinary Paradigm. *IEEE Wireless Communications*, *27*(4), 133–139. https://doi.org/10.1109/MWC.001.1900491
86. Klaine, P. V., Imran, M. A., Onireti, O., & Souza, R. D. (2017). A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks. In *IEEE Communications Surveys and Tutorials* (Vol. 19, Issue 4, pp. 2392–2431). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/COMST.2017.2727878
87. Chen, M., Challita, U., Saad, W., Yin, C., & Debbah, M. (2019). Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial. *IEEE Communications Surveys and Tutorials*, *21*(4), 3039–3071. https://doi.org/10.1109/COMST.2019.2926625
88. Hussain, F., Hassan, S. A., Hussain, R., & Hossain, E. (2020). Machine Learning for Resource Management in Cellular and IoT Networks: Potentials, Current Solutions, and Open Challenges. *IEEE Communications Surveys and Tutorials*, *22*(2), 1251–1275. https://doi.org/10.1109/COMST.2020.2964534
89. Awad Abdellatif, A., Abo-Eleneen, A., Mohamed, A., Erbad, A., Navkar, N. v., & Guizani, M. (2023). Intelligent-Slicing: An AI-Assisted Network Slicing Framework for 5G-and-Beyond Networks. *IEEE Transactions on Network and Service Management*, *20*(2), 1024–1039. https://doi.org/10.1109/TNSM.2023.3274236
90. Dahj, J. N. M., Ogudo, K. A., & Boonzaaier, L. (2024). A hybrid analytical concept to QoE index evaluation: Enhancing eMBB service detection in 5G SA networks. *Journal of Network and Computer Applications*, *221*. https://doi.org/10.1016/j.jnca.2023.103765
91. Kousias, K., Rajiullah, M., Caso, G., Alay, O., Brunstrom, A., Ali, U., de Nardis, L., Neri, M., & di Benedetto, M. G. (2024). Empirical performance analysis and ML-based modeling of 5G non-standalone networks. *Computer Networks*, *241*. https://doi.org/10.1016/j.comnet.2024.110207
92. Zyxel Networks, 2022, 5G optimal values, https://support.zyxel.eu/hc/en-us/articles/360005188999-Signal-quality-LTE-5G-LTE-and-5G-signal-quality-parameters#h_01H92W5VFVEF1PJMMNNQ2FM44G
93. 765G 5G Mobile Platform, prod_brief_qcom_sd765g_5g_1.pdf (qualcomm.com)
94. 480+ 5G mobile Platform, product_brief_-_snapdragon_480_plus_5g_mobile_platform.pdf (qualcomm.com)
95. Schober P, Boer C, Schwarte LA. Correlation Coefficients: Appropriate Use and Interpretation. Anesth Analg. 2018 May;126(5):1763-1768. doi: 10.1213/ANE.0000000000002864. PMID: 29481436. https://pubmed.ncbi.nlm.nih.gov/29481436/
96. 3GPP TS38.133, NR; Requirements for support of radio resource management
97. OSS Nokalva Inc., The new ISO 20022 financial messaging standard and its encodings, https://www.oss.com/asn1/resources/books-whitepapers-pubs/White%20paper%20on%20ASN%201%20in%20ISO%2020022%20-%2020130603.pdf