

Streszczenie

Rak szyjki macicy pozostaje jedną z głównych przyczyn zgonów kobiet na świecie, a jego wczesne wykrycie ma kluczowe znaczenie dla zmniejszenia śmiertelności. Jednak ręczna analiza obrazów cytologicznych szyjki macicy jest czasochłonna, subiektywna i w dużym stopniu zależna od doświadczenia patomorfologów, co często prowadzi do zmienności diagnoz. Z tego względu istnieje potrzeba opracowania zautomatyzowanych systemów komputerowego wspomagania diagnostyki (Computer-Aided Diagnosis, CAD), które umożliwią bardziej dokładne i spójne badania przesiewowe. W ostatnich latach głębokie uczenie (Deep Learning, DL) stało się skutecznym rozwiązaniem w analizie obrazów medycznych, ponieważ pozwala na automatyczne uczenie się złożonych wzorców wizualnych bez konieczności ręcznej ekstrakcji cech. W niniejszej pracy zaproponowano metody oparte na głębokim uczeniu do segmentacji i klasyfikacji komórek szyjki macicy z wykorzystaniem trzech zbiorów danych: prywatnego zbioru Pomeranian oraz publicznie dostępnych zbiorów SIPaKMeD i Herlev.

W zakresie segmentacji przeanalizowano kilka architektur DL, w tym Improved U-Net, Spatial Pyramid Pooling (SPP)-SegNet, squeeze-and-excitation DeepLabV3+ (SE-DeepLabV3+), Vision Transformer (ViT)-U-Net oraz model uncertainty-aware multi-tasking (UAMT). Improved U-Net rozszerza standardową architekturę U-Net poprzez udoskonalenie ekstrakcji cech oraz poprawę lokalizacji granic komórek szyjki macicy z wykorzystaniem współczynników dylatacji i normalizacji grupowej (Group Normalization). SPP-SegNet umożliwia uchwycenie wieloskalowych informacji kontekstowych, co poprawia segmentację komórek o zróżnicowanych rozmiarach. SE-DeepLabV3+ wykorzystuje bloki squeeze-and-excitation w celu lepszej rekaliibracji cech między kanałami oraz podkreślenia istotnych odpowiedzi cech. ViT-U-Net łączy warstwy konwolucyjne z blokami Vision Transformer, umożliwiając modelowanie zarówno lokalnych tekstur, jak i globalnych zależności przestrzennych struktur komórkowych. Ponadto model UAMT realizuje jednocześnie segmentację i klasyfikację oraz szacuje niepewność predykcji, zwiększając wiarygodność procesu diagnostycznego.

W zakresie klasyfikacji zbadano modele wykorzystujące mechanizmy uwagi (attention), uczenie zespołowe (ensemble learning) oraz hybrydowe architektury konwolucyjnych sieci neuronowych (CNN) i Vision Transformer (ViT), aby zwiększyć skuteczność i odporność modeli. Model oparty na mechanizmie uwagi adaptacyjnie koncentruje się na najbardziej istotnych obszarach komórek

szyjki macicy. Metoda residual-based ensemble łączy wiele niezależnie trenowanych, wstępnie wytrenowanych modeli CNN z mechanizmem rezyduального udoskonalania reprezentacji cech, co poprawia stabilność i dokładność predykcji. Hybrydowy model CNN–ViT integruje lokalną ekstrakcję cech realizowaną przez CNN z globalnym modelowaniem zależności charakterystycznym dla transformatorów. Dodatkowo zastosowano metodę learnable weight ensemble, która automatycznie uczy się optymalnych wag dla poszczególnych modeli podczas procesu treningu zamiast stosowania stałego uśredniania. W celu ograniczenia problemu niewielkiej liczby danych wykorzystano model Residual Deep Convolutional Generative Adversarial Network (RES-DCGAN) do augmentacji danych poprzez generowanie realistycznych syntetycznych obrazów komórek szyjki macicy, zwiększając różnorodność zbioru danych i odporność modeli.

Zaproponowane podejścia oceniono z wykorzystaniem podziału na zbiory treningowy i testowy, pięciokrotnej walidacji krzyżowej oraz walidacji międzyzbiorowej (cross-dataset validation). Uzyskane wyniki pokazują, że proponowane metody osiągają wysoką skuteczność na większości analizowanych zbiorów danych. W większości przypadków segmentacja zwiększa interpretowalność wyników poprzez precyzyjną lokalizację obszarów komórkowych. Jednocześnie wejścia oparte na obwiedniach (BBBox), obejmujące niewielkie marginesy wokół wysegmentowanych regionów, osiągają lepsze wyniki niż wejścia ROI (Region of Interest), ponieważ zachowują użyteczne informacje kontekstowe z otoczenia komórki. Modele wykorzystujące uczenie zespołowe oraz mechanizmy uwagi konsekwentnie przewyższają pojedyncze modele pod względem dokładności, stabilności oraz zdolności generalizacji.

Zaproponowane dwuetapowe modele segmentacji połączone z uczeniem zespołowym i klasyfikacją opartą na transformatorach mogą znacząco poprawić skuteczność wykrywania raka szyjki macicy. Ponadto opracowany model dostarcza użytecznych wyjaśnień wizualnych oraz wiarygodnego wsparcia decyzyjnego, co czyni go odpowiednim do zastosowań w rzeczywistych systemach CAD. Uzyskane wyniki są konkurencyjne, a w wielu przypadkach przewyższają rezultaty prezentowane w dotychczasowych badaniach, potwierdzając skuteczność zaproponowanych metod.

Słowa kluczowe: Głębokie uczenie, segmentacja obrazów, klasyfikacja obrazów, ulepszony U-Net, ensemble rezydualny, Vision Transformer (ViT), augmentacja danych, rak szyjki macicy.