Rzeszów, 03 października 2026 r.

prof. dr hab. inż. Mariusz Oszust
Katedra Informatyki i Automatyki
Wydział Elektrotechnik i Informatyki
Politechnika Rzeszowska

Recenzja rozprawy doktorskiej mgr. inż. Michała Wieczorka pt. „Semi-supervised Training with Fuzzy-Logic for Transformer-Based Semantic Segmentation”

Recenzja została przygotowana na zlecenie przewodniczącego Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja Politechniki Częstochowskiej, dra hab. inż. Adama Kulawika, prof. PCz (Uchwała nr 29/2025/2026 z dnia 25.06.2026 r., pismo R-WIiSI-BOD-48-2026 dra hab. inż. Adama Kulawika, prof. PCz z dnia 29.06.2026 r., pismo R-WIiSI-BOD-52-2026 dr inż. Aleksandry Krzywik z dnia 03.08.2026 r.). Postępowanie jest prowadzone w dziedzinie nauk inżynieryjno-technicznych, w dyscyplinie Informatyka Techniczna i Telekomunikacja. Promotorem rozprawy jest prof. dr hab. inż. Marcin Woźniak.

Zgodnie z danymi udostępnionymi w systemie informacji o nauce i szkolnictwie wyższym RAD-on Doktorant jest absolwentem Politechniki Śląskiej, na której ukończył studia na kierunku informatyka, uzyskując tytuł zawodowy inżyniera w 2022 roku, a następnie magistra inżyniera w roku 2024. Od 1 marca 2024 roku Doktorant jest zatrudniony w macierzystej jednostce, która stanowi jego podstawowe miejsce pracy, przy czym jako wiodącą dyscyplinę naukową zadeklarował Informatykę Techniczną i Telekomunikację.

Recenzowana rozprawa doktorska podejmuje niezwykle aktualną tematykę projektowania dedykowanych architektur głębokich sieci neuronowych przeznaczonych do segmentacji semantycznej obrazów medycznych, ze szczególnym uwzględnieniem preparatów hematologicznych. Zakres pracy wykracza jednak znacznie poza samą propozycję nowych architektur, ponieważ tworzy oprogramowanie z myślą o barierach współczesnej analizy obrazów medycznych - problematyce niedoboru danych uczących oraz wyzwaniach związanych z niejednoznacznością adnotacji eksperckich. Dopełnieniem tych rozważań jest

aspekt aplikacyjny, skupiający się na architekturach uczenia federacyjnego oraz konstruowaniu bezpiecznych, odpornych na ataki systemów wspomagania decyzji.

Tematykę rozprawy uważam za aktualną istotną z punktu widzenia rozwoju dyscypliny Informatyka Techniczna i Telekomunikacja. Zasadnicza część rozprawy jest napisana w języku angielskim i obejmuje 146 stron. Całość składa się ze streszczeń w języku polskim i angielskim, przedstawienia tez, 11 rozdziałów i spisu literatury zawierającego 142 pozycji.

W rozdziale poświęconym tezom Doktorant (niejawnie) zdefiniował główny cel pracy jako opracowanie opartego na sieci Transformer półnadzorowanego środowiska programistycznego – szkieletu do segmentacji semantycznej obrazów mikroskopowych rozmazów krwi w hematologii. Rozwiązanie ma uwzględniać ograniczenia wynikające z braku adnotacji, zmienności barwienia preparatów, nakładania się komórek na siebie oraz niejednoznaczności ich krawędzi. Następnie, na stronie iii, Doktorant zdefiniował trzy tezy związane z (1) rozwojem mechanizmu uwagi ukierunkowanym na poprawę zdolności generalizacji metody semantycznej segmentacji opartej na Transformerze, (2) rozmytym i półnadzorowanym algorytmem uczenia do celów semantycznej segmentacji poprawiającym klasyfikację na poziomie pikseli przy niejednoznacznych danych, oraz (3) wprowadzeniem technik rozmytych do architektur semantycznej segmentacji by zredukować błędy klasyfikacji na poziomie pikseli przy niejednoznacznych danych i granicach obiektów. Poniżej tez Doktorant przedstawia środowisko programistyczne, szerzej opisane w rozdziale 9, które zbudował korzystając z doświadczeń zdobytych podczas rozwiązywania problemów cząstkowych, przedstawionych we wcześniejszych rozdziałach.

Rozdział pierwszy krótko wprowadza do zagadnień związanych ze sztuczną inteligencją, skupiając uwagę na jej zastosowaniach w medycynie oraz potrzebie klasyfikacji tkanek na zdjęciach mikroskopowych. Doktorant podkreśla, że znane rozwiązania wykorzystujące głębokie sieci neuronowe posiadają ograniczenia związane z zawężonym obszarem obserwacji. Ograniczenia te są niwelowane dzięki pojawieniu się architektur opartych na sieciach Transformer, które dzięki wykorzystaniu mechanizmu uwagi osiągają lepsze rezultaty, zwłaszcza że potrafią uwzględniać zależności długoterminowe oraz globalny kontekst. Przechodząc do segmentacji semantycznej obrazów hematologicznych, naświetlony zostaje problem niejednoznaczności danych medycznych, w przypadku których często niemożliwe

jest precyzyjne określenie położenia komórki czy jej granic ze względu na położenie innych komórek. Z tego powodu zastosowanie podejść opartych na logice rozmytej, stosowanych w tym przypadku w funkcji straty lub do określania pseudo-etykiet dla danych nieoznaczonych, wydaje się uzasadnione.

Rozdział drugi zawiera wykaz terminów i określeń z zakresu sztucznej inteligencji, wizji komputerowej oraz analizy obrazów medycznych. Taki słownik pojęć stanowi wartościową pomoc dla czytelników mniej zaznajomionych z tematyką rozprawy.

W rozdziale trzecim Doktorant zwięźle wprowadza podstawy teoretyczne dotyczące sieci neuronowych i mechanizmu uwagi, przedstawia architekturę Transformer oraz metody optymalizacji i regularyzacji.

Rozdział czwarty dotyczy detekcji malarii w próbkach krwi i stanowi etap wstępny dla głównego nurtu badań rozprawy. W podrozdziale 4.1 wprowadzono problematykę detekcji malarii, znaczenie przeciwdziałania tej chorobie oraz jej wpływ na populację, a także przedstawiono dotychczasowe metody jej wykrywania. Po omówieniu rozwiązań opartych na sieciach neuronowych, Doktorant wskazuje na wyzwania związane z koniecznością znalezienia kompromisu pomiędzy skutecznością detekcji, złożonością modeli a wymogiem precyzyjnej segmentacji komórek, nakreślając sposób podejścia do tych problemów w autorskim rozwiązaniu. Podrozdział 4.2 dodatkowo podkreśla zapotrzebowanie na nowe narzędzia diagnostyczne, w szczególności metody szybkie, dokładne i wymagające tylko pojedynczego badania. W podrozdziale 4.3 Doktorant przedstawia podstawy uczenia głębokiego i segmentacji semantycznej w ujęciu aplikacyjnym dla detekcji malarii (odwołując się m.in. do fragmentów 3.5-3.9), natomiast w podrozdziale 4.4 opisuje proces przygotowania zbioru danych. Na bazie publicznie dostępnego zbioru, przy użyciu zaproponowanego algorytmu transformacji zgrubnych adnotacji z ramką ograniczającą (ang. bounding-box) do segmentacji na poziomie pikseli, Autor wygenerował podzbiór 1364 obrazów (obejmujący około 80 tys. komórek) na potrzeby weryfikacji własnego modelu. Następnie zaprojektował sieć opartą na symetrycznej strukturze typu enkoder-dekoder z blokami rezydualnymi oraz połączeniami charakterystycznymi dla architektury U-Net. Architektura i proces uczenia powstałej sieci opisał w rozdziale 4.5. W rozdziale 4.6 przedstawiono wyniki detekcji wyrażone poprzez skuteczność segmentacji na poziomie pikseli, a także zestawienie tych rezultatów z innymi pracami. Całość uzupełnia wizualizacja próbek

testowych, na których wyodrębniono maski tła oraz komórki zdrowe i zainfekowane. Krótkie podsumowanie znajduje się w rozdziale 4.8.

O ile w rozdziale czwartym Doktorant opisywał rozwiązanie oparte na klasycznej już sieci splotowej, o tyle rozdział piąty prezentuje podejście wykorzystujące architekturę Transformer. Celem było zweryfikowanie, czy globalna samouwaga (ang. self-attention) jest w stanie dostarczyć kontekstu globalnego, który pozostaje niedostępny dla architektur splotowych. W podrozdziale 5.1 Autor uszczegóławia ten problem, wskazując na konieczność uzyskania precyzyjnej segmentacji, umożliwiającej dokładną analizę struktur cienkich lub trudnych do wyodrębnienia. Aby właściwie umiejscowić zaproponowaną metodę, w podrozdziale 5.2 dokonano przeglądu prac pokrewnych i zdiagnozowano ich ograniczenia. Doktorant uzasadnia tym samym potrzebę wdrożenia architektur typu Transformer w dziedzinie dotychczas zdominowanej przez klasyczne sieci splotowe. W podrozdziale 5.3 wprowadzono podstawy teoretyczne mechanizmów uwagi oraz architektury Transformer i sposobów jej adaptacji. Wyszczególniono przy tym trzy znane strategie oparte na: klasyfikacji wycinków obrazu, wykorzystaniu Transformera jako metody wyodrębniania cech oraz zastosowaniu Transformera z mechanizmem uwagi krzyżowej zachodzącej pomiędzy wyjściem enkodera a wyuczonymi osadzeniami. Rozwiązanie zaproponowane przez Doktoranta należy do drugiej z wymienionych grup. Wprowadził on jednak autorskie modyfikacje w obrębie samego enkodera oraz interfejsu łączącego enkoder z dekoderem.

W celu przetestowania zaproponowanej architektury Doktorant wykorzystał zbiór 366 obrazów obejmujący trzy klasy: tło, krętki (Spirochaete) oraz erytrocyty (Erythrocyte). Zastosowano również wzbogacanie zbioru treningowego, wykorzystując zmianę rozmiaru obrazu, przycinanie, obracanie oraz transformacje fotometryczne. Zaproponowana architektura została poddana ewaluacji, a jej skuteczność segmentacji zaraportowano z użyciem pięciu metryk (mIoU, mDice, czułość, precyzja oraz współczynnik F1). W podrozdziałach 5.5–5.7 ocenę ilościową uzupełniono o wizualizacje oraz testy ablacyjne. Badania te w pełni uzasadniają wprowadzone modyfikacje względem klasycznej architektury ViT i potwierdzają poprawę wyników w stosunku do klasycznych sieci splotowych. W szczególności wykazano zasadność zastosowania enkodera hierarchicznego zamiast płaskiego, połączenia go z dekoderem splotowym z wykorzystaniem połączeń znanych z sieci U-Net, a także

wykorzystania osadzeń wycinków z kodowaniem pozycyjnym. Uzyskane rezultaty odniesiono do innych metod z literatury przedmiotu. Z kolei podrozdziały 5.8 i 5.9 opisują ograniczenia metody i podsumowują zaprezentowane rozwiązanie, wskazując na zalety modyfikacji dopasowanych do wymagań segmentacji semantycznej.

Rozdział szósty dotyczy segmentacji semantycznej zastosowanej do problemu detekcji broni. Ze względu na niedostatek odpowiednio adnotowanych zbiorów, zaproponowane podejście opiera się na wykorzystaniu sztucznie generowanych danych. Problematyka i waga automatycznych metod wykrywania broni zostały nakreślone w podrozdziale 6.1, gdzie wyraźnie wskazano na braki w dostępności odpowiednich zbiorów uczących. Z kolei w podrozdziale 6.2 Doktorant dokonuje krótkiego przeglądu literatury i dostępnych zbiorów syntetycznych. Podrozdział 6.3 stanowi wprowadzenie do generowania danych syntetycznych. Autor akcentuje w nim problemy klas rzadkich wynikające z nietypowych kątów obserwacji lub mało znanych modeli broni oraz okluzji. Trafnie diagnozuje trudności związane z wykorzystaniem danych rzeczywistych, takie jak wysoki koszt i niedostępność, brak kontroli nad rozkładem klas, kwestie etyczne i prawne oraz problem wierności adnotacji, wymagającej żmudnego śledzenia konturów obiektów. Wykazano, że metody oparte na danych syntetycznych stanowią atrakcyjną alternatywę pozbawioną tych ograniczeń. Podrozdział 6.4 zawiera szczegółowy opis algorytmu syntezy obrazów, uwzględniającego m.in. modyfikacje oświetlenia, zmianę tła, dodawanie okluzji oraz symulację parametrów optycznych, takich jak ogniskowa obiektywu i rozmycie. Następnie w podrozdziale zaprezentowano 66-warstwową sieć splotową o symetrycznej architekturze enkoder-dekoder. Kolejne podrozdziały (6.5 - 6.6) opisują proces uczenia sieci oraz uzyskane rezultaty detekcji na tle analogicznych metod z literatury. Wartościowym elementem tej części jest analiza tzw. trudnych przypadków, w których model błędnie klasyfikował próbki. Dotyczyło to głównie silnego rozmycia, nietypowych konstrukcji broni, cieni, prześwietleń oraz wpływu kontekstu wizyjnego. Rozdziały 6.7 oraz 6.8 poświęcono dyskusji nad ograniczeniami metody. Wskazano tu na problem danych generowanych sztucznie, które nie w pełni oddają rzeczywiste zniekształcenia lub specyficzne czynniki środowiskowe (np. deszcz, czy specyfikę szumu z innych kamer). Rozdział zamyka

podsumowanie akcentujące potencjalny, pozytywny wpływ tego typu rozwiązań na bezpieczeństwo publiczne.

Rozdział siódmy przedstawia nieco inne podejście, związane z uczeniem federacyjnym, jednak nadal koncentruje się na problematyce niepewności zawartej w próbkach danych do segmentacji. W podrozdziałach 7.1 i 7.2 wprowadzono koncepcję uczenia federacyjnego, zaprezentowano dotychczasowe dokonania Doktoranta w tym zakresie oraz omówiono architekturę tego paradygmatu, uwzględniając zagrożenia bezpieczeństwa i wektory ataków. Zamieszczono tam również zbiorcze zestawienie takich ataków wraz ze zwięzłym przeglądem metod ochrony i ich ograniczeń. Podrozdział 7.3 zawiera uzasadnienie dla wykorzystania logiki rozmytej drugiego rzędu (typu 2). Jak słusznie zauważa Autor, klasyczne zbiory niosą jedynie binarną informację o przynależności do klasy, a zbiory rozmyte pierwszego rzędu modelują przynależność częściową; dopiero zastosowanie zbiorów drugiego rzędu pozwala na uwzględnienie dodatkowej informacji o niepewności. Architekturę systemu wnioskowania oraz operacje na tego typu zbiorach przedstawiono w sposób opisowy, wyodrębniając pięć etapów: rozmywanie, ewaluację lingwistycznej bazy reguł, agregację, redukcję typu oraz wyostrzanie. Poszczególne komponenty tej architektury, wraz z teoretyczną analizą mechanizmu reputacji, zawarto w podrozdziale 7.4. Zaproponowane podejście zostało przetestowane w różnych scenariuszach ataku i porównane z analogicznymi metodami, osiągając zadowalające rezultaty (podrozdział 7.5). Wykonane prace podsumowano w podrozdziale 7.6, w którym Doktorant podkreślił spójność tego podejścia z pozostałymi rozwiązaniami proponowanymi w rozprawie, ponownie nawiązując do fundamentalnego problemu niejednoznaczności danych medycznych.

Chcąc przetestować wpływ własnej koncepcji architektury oraz generowania syntetycznych danych na odporność opracowanego rozwiązania, Doktorant utworzył automatyczny potok przetwarzania. System ten pobiera obrazy mikroskopowe w jasnym polu, przedstawiające skrawki tkankowe wybarwione hematoksyliną i eozyną (H&E), poddaje je augmentacji, trenuje sieć neuronową i ostatecznie zwraca rozkłady prawdopodobieństwa dla 25 klas narządów i gatunków. Rozwiązanie to opisano w rozdziale ósmym. Podrozdział 8.1 stanowi wstęp do zagadnienia, natomiast 8.2 przedstawia proces przygotowania danych, zrealizowany we współpracy z Zakładem Histologii i Embriologii Uniwersytetu Przyrodniczego

w Lublinie. Zbiór objął pośmiertne próbki tkanek bydła, świń i koni (stanowiące produkty uboczne z rzeźni) oraz dane historyczne dotyczące preparatów ptasich (m.in. żołądków mięśniowych). Ostatecznie uzyskano 3680 obrazów pokrywających 25 specyficznych kombinacji narząd-gatunek. Po etapie wstępnego przetwarzania danych (opisanym w podrozdziale 8.3), zbiór poddano augmentacji (podrozdział 8.4) obejmującej rotację, odbicia lustrzane oraz transformacje fotometryczne, w tym zmianę przestrzeni barw na HSL i zaburzenia kolorów. Ciekawym i wartościowym elementem tej pracy jest to, że operacje te były sprawdzane statystycznie. Z wykorzystaniem dywergencji Kullbacka-Leiblera weryfikowano czy wygenerowane obrazy charakteryzują się rozkładem zbliżonym do danych oryginalnych, a za pomocą testu Kołmogorowa-Smirnowa badano czy są one statystycznie wierne. Architektura systemu, oparta na sieci ResNet, została szczegółowo opisana w podrozdziałach 8.5 i 8.6. Wyniki na zebranym zbiorze przedstawiono w podrozdziale 8.7, raportując pięć miar skuteczności. Rozdział zawiera również analizę macierzy pomyłek, która potwierdza wysoką skuteczność zaproponowanego rozwiązania. Z kolei podrozdziały 8.8 i 8.9 stanowią dyskusję nad wyborem rozwiązań cząstkowych i uzasadnienie doboru sieci i funkcji straty, a także potencjalnych źródeł błędu systematycznego. Całość zamyka podsumowanie, które uwypukla główne atuty metody: znakomitą skuteczność (ponad 98%) przy zachowaniu niskiego zapotrzebowania na moc obliczeniową dzięki wykorzystaniu płytkiej sieci neuronowej.

Rozdział dziewiąty prezentuje docelowy szkielet półnadzorowanej metody segmentacji semantycznej obrazów hematologicznych. Podrozdział 9.1 stanowi ponowne wprowadzenie do tematyki, tym razem z perspektywy zastosowania sieci typu Transformer oraz wbudowania mechanizmów uczenia półnadzorowanego, rozmytego nadzoru i regularyzacji bezpośrednio w funkcję straty. Jak zaznaczono w podrozdziale 9.2, zaproponowane rozwiązanie stanowi finalną aplikację, będącą syntezą doświadczeń i metod wypracowanych na wcześniejszych etapach badawczych ujętych w rozprawie. Zasadność budowy tego typu systemów wspomagania decyzji zarysowano w sekcji 9.3 i poparto przeglądem literatury (podrozdział 9.4). W podrozdziale 9.5 sformułowano problem badawczy, opierając się na podziale zbioru na dane posiadające adnotacje oraz ich pozbawione. Głównym celem optymalizacji jest estymacja parametrów modelu poprzez złożoną

funkcję straty. Składa się ona z dwóch oddziałujących na siebie członów: rozmytego grupowania z regularyzacją dla danych bez etykiet oraz standardowej straty segmentacji dla danych zaetykietowanych. W efekcie zoptymalizowano system tak, aby dane bez etykiet modelowały ogólną strukturę obrazu, natomiast te z etykietami odpowiadały za prawidłowe przypisanie semantyki. Architekturę sieci zdefiniowano w podrozdziale 9.6. Doktorant zaadaptował tu mechanizmy znane z architektury SegFormer, wykorzystując hybrydowy mechanizm globalnej uwagi wspierany pamięcią typu pooling oraz kompresją spektralną. Ocenę eksperymentalną rozwiązania przedstawiono w podrozdziale 9.7. Do weryfikacji rezultatów, obok standardowej dokładności walidacji, Doktorant wykorzystał miary micro-Dice oraz micro-IoU. Podejście to jest szczególnie pożądane przy analizie globalnej skuteczności detekcji na zbiorach o niezrównoważonym rozkładzie klas. Ilościową ewaluację dopełniają wizualizacje potwierdzające zgodność wygenerowanych masek z referencyjnymi adnotacjami eksperckimi. Porównanie z innymi pracami (podrozdział 9.8) przeprowadzono na podstawie wyników raportowanych w literaturze, bez bezpośredniej reprodukcji wyników dla modeli na wspólnym zbiorze uczącym. Uzyskane rezultaty okazały się jednak konkurencyjne, co Autor trafnie argumentuje przewagami wynikającymi ze specyfiki zaprojektowanej architektury sieci. Całość zamyka rozdział 9.9, który podsumowuje badania i jednoznacznie wskazuje na duży potencjał aplikacyjny architektur Transformer w zautomatyzowanej segmentacji semantycznej oraz analizie rozmazów krwi.

Rozdział dziesiąty stanowi krótką, zbiorczą dyskusję nad rozwiązaniami przedstawionymi w poprzednich rozdziałach, opatrzoną dodatkowymi uzasadnieniami. Z kolei rozdział jedenasty to podsumowanie całej rozprawy.

Doktorant realizując cel badawczego uzyskał szereg nowych rezultatów. Do najważniejszych zaliczam:

- Opracowanie metody konwersji zgrubnych adnotacji ramkowych do masek segmentacyjnych na poziomie pikseli, zastosowanej do wygenerowania referencyjnego zbioru danych w zadaniu detekcji malarii.
- Zaprojektowanie sieci do analizy obrazów medycznych, łączącej hierarchiczny enkoder, posiadający mechanizm globalnej samouwagi i kodowanie pozycyjne, z dekoderem splotowym.

- Stworzenie zautomatyzowanego potoku syntezy danych obrazowych, uwzględniającego modyfikacje oświetlenia, tła, okluzje oraz parametry optyczne na potrzeby trenowania sieci.
- Zastosowanie rozmytych systemów wnioskowania typu 2. do modelowania niepewności w rozproszonych danych segmentacyjnych wraz z mechanizmem reputacji chroniącym system przed atakami.
- Opracowanie docelowej metody bazującej na sieci SegFormer z hybrydowym mechanizmem uwagi, wykorzystującej nową funkcję straty, która integruje nadzorowaną stratę segmentacji dla danych zaetykietowanych z rozmytym grupowaniem dla danych bez etykiet.

Do zasadniczych słabości rozprawy należą częste uogólnienia, brak rozszerzonej dyskusji ze względu na decyzję o przedstawieniu wielu różnych rozwiązań, niedostateczna precyzja sformułowań oraz powierzchowna prezentacja weryfikacji eksperymentalnej uzasadniająca wyciągane wnioski.

Uwagi merytoryczne:

1) Wykaz terminów i określeń z zakresu sztucznej inteligencji zajmuje 12 stron i być może byłby przydatny osobom niezaznajomionym z tematyką rozprawy, jednak trudno uznać, aby lakoniczny sposób przedstawiania często trudnych pojęć spełniał swoją rolę. Znacznie szczęśliwszym rozwiązaniem byłoby przygotowanie solidnego wprowadzenia do tematu, rozszerzającego obecne wstępy do rozdziałów. Wówczas tłumaczenie czym jest „piksel” byłoby zbędne, a opis reguł rozmytych nie sprowadzałby się do intuicji, lecz zawierałby ścisłe informacje o zbiorach rozmytych, funkcjach przynależności, zmiennych, operatorach, wnioskowaniu, wyostrzaniu, czy typach reguł. Stwierdzenie, że reguły sprawiają, iż systemy rozmyte są „intuicyjne i ekspresywne”, jest w tym miejscu nadmiarowe, zwłaszcza że są to określenia subiektywne a większość pojęć w słowniku nie zawiera tego typu podsumowań.
2) W rozprawie zabrakło osobnego rozdziału lub podrozdziału z metrykami, a sposób ich wyznaczania znajduje się dopiero na stronie 57 (rozdział 6.6).
3) Tabela 1 na stronie 30 zawiera porównanie metod do detekcji malarii, którego analiza znajduje się na stronie 29. Stwierdzenie, że zaproponowany model

przewyższa inne podejścia, ponieważ osiągnął 99,68% skuteczności detekcji na własnym zbiorze, nie jest w pełni uprawnione. Inne metody były bowiem testowane na odmiennych danych – dla przykładu, te z pozycji [56] i [57] walidowano na znacznie większym zbiorze niż ten użyty w rozprawie. O ile podobne porównania w późniejszych rozdziałach są odpowiednio opatrzone komentarzem, a wyciągane wnioski co do przewagi własnych rozwiązań są znacznie bardziej stonowane, o tyle w tym rozdziale wyraźnie tego zabrakło. Przy tego typu zestawieniach zawsze warto wyjaśnić, jakiemu celowi służy porównanie, oraz jawnie wskazać zbiory, na których uzyskano poszczególne wyniki, aby uniknąć formułowania pochopnych wniosków.

4) W pracy brakuje rysunków przedstawiających schematy sieci neuronowych. Uwaga ta dotyczy w szczególności architektur zaproponowanych przez Autora. Zamieszczenie takich schematów, wraz z wyraźnym zaznaczeniem nowych elementów wprowadzonych w rozprawie, znacznie ułatwiłoby ocenę rzeczywistego wkładu proponowanych modeli w rozwój dyscypliny.
5) Porównanie w Tabeli 1 opiera się na wskaźniku skuteczności detekcji. Czy stosowanie tej miary jest odpowiednie w przypadku walidacji metody na danych niezbilansowanych? Do jakich błędnych interpretacji może to doprowadzić i jakie metryki powinny być tutaj docelowo zastosowane?
6) Dyskutując ograniczenia metody na stronie 32, Autor stwierdza, że jest ona odporna na wiele artefaktów. Rozdział nie zawiera jednak opisów eksperymentów, które mogłyby tę tezę potwierdzić. Brakuje także doprecyzowania o jakich artefaktach mowa oraz jaka część analizowanego zbioru je zawiera.
7) W Algorytmie 1 zaprezentowano procedurę generowania danych syntetycznych. W tekście zabrakło jednak szerszego uzasadnienia dotyczącego kryteriów wyboru zastosowanych metod oraz doboru ich parametrów. W jakim stopniu uzyskane w ten sposób obrazy przypominają dane rzeczywiste, zwłaszcza po zaaplikowaniu transformacji fotometrycznych i losowych modyfikacji wartości pikseli (ColorJitter)?
8) Opisane w rozdziale szóstym generowanie danych syntetycznych stanowi główny wkład badawczy Autora. Choć Doktorant zastosował szereg zaawansowanych metod w celu uzyskania wiarygodnych obrazów, w pracy nie zaprezentowano żadnych wizualnych przykładów wygenerowanych w ten

sposób danych. Wartościowe byłoby również zilustrowanie omawianych na stronie 60 tzw. przypadków trudnych, np. referowane w punkcie nr 4 podobieństwo wiertarki do broni powinno być poparte zdjęciem.

9) Zawarta na stronie 70 analiza teoretyczna cech mechanizmu reputacji wydaje się zbyt powierzchowna i wymagająca dodatkowych uzasadnień. Znacznie lepszym rozwiązaniem byłoby praktyczne wykazanie na konkretnych przykładach, w jaki sposób system rozmyty poprawia działanie tego mechanizmu w uczeniu federacyjnym. Decydując się jednak na ujęcie formalne, Doktorant powinien nanieść następujące poprawki: (i) Wyraźnie uzasadnić, dlaczego do analizy wybrano akurat te trzy konkretne właściwości. (ii) Doprecyzować i rozszerzyć opisy. Przykładowo w Twierdzeniu 1 zmienne D_i oraz D_j wymagają zdefiniowania. Ponadto, zamiast jedynie deklarować istnienie dowodu, należy go rzetelnie przeprowadzić.
W przeciwnym razie twierdzenia powinny zostać nazwane np. założeniami. (iii) Usunąć lub istotnie przeredagować Twierdzenie 3 -stwierdzenie, że sam brak dostępu do surowych danych gwarantuje pełną prywatność jest zbyt definitywne. W systemach federacyjnych udostępniane wyniki mogą być na tyle specyficzne dla danego klienta, że pozwalają napastnikowi na odtworzenie wrażliwych informacji o charakterze jego danych treningowych.
10) Pewne wątpliwości budzi stwierdzenie ze strony 107, gdzie Autor deklaruje, że pomimo ograniczonej dostępności odpowiednio przygotowanych przykładów, zaproponowany system jest dobrze dopasowany do rozwiązywanego problemu. Czy korpus liczący zaledwie 100 obrazów można uznać za wystarczający do walidacji modelu i wyciągania tak kategorycznych wniosków? W pracy zabrakło określenia, jaki docelowy rozmiar powinien mieć zbiór danych, aby można było mówić o rzetelnym i wszechstronnym przetestowaniu zaproponowanego podejścia. Czy w literaturze lub publicznych repozytoriach istnieją inne, alternatywne zbiory, które można by w tym przypadku zaadaptować?

Mniej istotne, wybrane usterki redakcyjne:

- Streszczenie wykonane w języku polskim zawiera kilka nietypowych i wprowadzonych przez Doktoranta określeń, które, porównując tekst w języku angielskim, mogą utrudniać zrozumienie intencji. O ile adaptacja metod do

nowych zastosowań jest powszechnie znana i stosowana, to „pryncypialne strategie efektywności danych… w zastosowaniach niskodanych” wymagają dodatkowych wyjaśnień.

- Słowa kluczowe w streszczeniu w języku polskim zostały zapisane bez polskich znaków.
- Strona iii. Literówka „details om”.
- Niektóre terminy są wielokrotnie tłumaczone i objaśniane w treści pracy (np. AI, ViT, CNN). Zgodnie z dobrą praktyką redaktorską, zasadne jest utworzenie wykazu skrótów i konsekwentne posługiwanie się nimi w tekście, bez ponownego ich rozwijania. Ponadto zaprezentowany słownik pojęć jest nadmiernie rozległy, trudno określić czy wszystkie sformułowania są istotne i wykorzystywane w rozprawie (np. FFT).
- Rozdział nr 3 zawiera teoretyczne wprowadzenie, które jest redundantne biorąc pod uwagę później przedstawiane treści. Dla przykładu mechanizm uwagi - wzór (5) w czterozdaniowym podrozdziale 3.6 można znaleźć w podrozdziale 5.3. Przy okazji w 3.6 mamy zapis z uwzględnieniem notacji macierzowej, zaś w 5.3 z niej zrezygnowano. Początek rozdziału 9.1 również zdaje się powielać niektóre treści z poprzednich rozdziałów.
- Niektóre tabele mają podpis nad tabelą, a inne pod tabelą.
- W tabeli 13 w kolumnie „No attack” pogrubiono gorszy wynik.
- Strona 37. Brakuje przecinków lub kropek po wzorach.
- W części pracy Autor stosuje „x” zamiast „\times” przy podawaniu rozdzielczości obrazów. W późniejszych rozdziałach jest poprawnie.

Wskazane w powyższej części uwagi, poza drobnymi kwestiami redakcyjnymi, mają charakter polemiczny i nie podważają merytorycznej wartości recenzowanej rozprawy. Warto zaznaczyć, że wyznaczony cel pracy został zrealizowany, a uzyskane rezultaty stanowią potwierdzenie słuszności tez badawczych. Rozprawa jest merytorycznie spójna i zgodna ze standardami stawianymi pracom doktorskim, oferując zarówno przegląd stanu wiedzy, jak i ewaluację nowych metod. Praca potwierdza samodzielność naukową Doktoranta, jego szeroką wiedzę w zakresie podejmowanej tematyki oraz biegłość w posługiwaniu się warsztatem badawczym.

Biorąc pod uwagę powyższe oraz wymagania określone w art. 187 ustawy Prawo o szkolnictwie wyższym i nauce z dnia 20 lipca 2018 r. (t.j. Dz. U. z 2024 r. poz. 1571 z późn. zm.), stwierdzam, że rozprawa doktorska mgr. inż. Michała Wieczorka zawiera oryginalne rozwiązanie problemu naukowego, a Kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka Techniczna i Telekomunikacja oraz umiejętność samodzielnego prowadzenia pracy naukowej.

W związku z powyższym pozytywnie opiniuję przedstawioną rozprawę doktorską, wnoszę o jej przyjęcie oraz dopuszczenie jej Autora, mgr. inż. Michała Wieczorka, do dalszych etapów postępowania, w tym do publicznej obrony rozprawy.