
Recenzja rozprawy doktorskiej

Tytuł rozprawy: Role of Spatial and Temporal Context in Classification Using Deep Neural Networks

Autorka: Mgr inż. Paulina Anna Tomaszewska

Promotor: Prof. dr hab. inż. Przemysław Biecek

1 Tematyka rozprawy

Recenzowana rozprawa doktorska Pani mgr inż. Pauliny Tomaszewskiej dotyczy zagadnień związanych z analizą wykorzystania informacji kontekstowej, np. przestrzennej lub czasowej, w systemach sztucznej inteligencji, zwłaszcza opartych na uczeniu głębokim, ze szczególnym akcentem położonym na metody wyjaśnialnej sztucznej inteligencji (z ang. *explainable artificial intelligence*, XAI). Doktorantka zauważa, że obecnie stosowane techniki XAI nie uwzględniają relacji pomiędzy obiektami w danych obrazowych, często ograniczając się do weryfikacji roli pojedynczych obiektów w procesie podejmowania decyzji. Jeśli informacje związane z kontekstem, np. przestrzennym, są istotne, to – jak wskazuje Doktorantka – modele głębokie nie mogą być właściwie ewaluowane. Praca doktorska Pani mgr inż. Pauliny Tomaszewskiej jest istotnym krokiem w stronę zaadresowania tej luki badawczej.

Uważam, że tematyka rozprawy jest ciekawa i niezwykle aktualna. Tematyka jest zgodna z obecnym nurtem badań związanych z tworzeniem wyjaśnialnych systemów sztucznej inteligencji oraz z projektowaniem technik XAI, które mogłyby pozwolić na wszechstronne zweryfikowanie opracowywanych algorytmów. Warto podkreślić, że poprawne wykorzystanie metod wyjaśnialnej sztucznej inteligencji może nie tylko pozwolić na ich łatwiejsze wdrożenie w praktycznych zastosowaniach (np. w medycynie czy w analizie danych satelitarnych, tj. w dziedzinach, w których wyniki działania modelu i sposób, w jaki zostały one wypracowane, powinny być interpretowalne), ale także na opracowanie ulepszonych algorytmów AI. Podjęta tematyka dotyczy nie tylko bardzo istotnych aspektów teoretycznych, związanych z analizą wpływu i istotności informacji kontekstowych w systemach sztucznej inteligencji, ale także jest niezwykle istotna z praktycznego punktu widzenia w szeregu zastosowań związanych z wizją komputerową, analizą danych obrazowych (np. w medycynie) czy analizą szeregów czasowych (np. w detekcji anomalii w danych satelitarnych).

2 Ocena strony merytorycznej rozprawy doktorskiej

Rozprawa doktorska jest napisana w języku angielskim i składa się z trzynastu numerowanych rozdziałów, bibliografii, spisu rysunków i tabel. **Rozdział pierwszy** jest wprowadzeniem do tematyki rozprawy – po krótkim (ale treściwym) wprowadzeniu, Autorka uwypukla najistotniejszą lukę badawczą, którą chce zaadresować w ramach niniejszej pracy (tj. pomijanie informacji i zależności kontekstowych w technikach XAI). Doktorantka płynnie, w sekcji 1.1, przechodzi do przedstawienia poprawnie sformułowanej hipotezy i dokładnie omawia aspekty, które będą kluczowe w badaniach. Bardzo dobrym pomysłem było przejrzyste pokazanie głównych osi badań, w ramach których będzie poruszać się Doktorantka – dotyczą one rodzaju informacji kontekstowej, rozmiaru tego kontekstu, konkretnych zastosowań i tego, czy kontekst jest analizowany na etapie tworzenia czy oceny algorytmów uczenia głębokiego (w kolejnych rozdziałach Doktorantka będzie „podświetlać” aspekty, które dotyczą konkretnych badań i opracowywanych metod). W kolejnej sekcji 1.2 Autorka dekomponuje hipotezę badawczą w 7 precyzyjnie postawionych pytań badawczych (RQ1–7) – dla każdego z pytań wskazano, w której z kluczowych (w kontekście rozprawy) publikacji Doktorantki [P1]–[P8] dane zagadnienie zostało podjęte. Na Rys. 1.1 Doktorantka prezentuje, w jaki sposób wspomniane cztery aspekty badań zostały podjęte w każdej z prac – warto podkreślić, że takie graficzne podsumowanie bardzo ułatwia szybkie zrozumienie tych prac. Rysunek 1.2 jest wizualnym podsumowaniem najważniejszych wyników prac Doktorantki. Być może warto byłoby uwypuklić, np. kolorem, które z przedstawionych bloków dotyczą nowych metod, wyników badań/obserwacji eksperymentalnych oraz danych? W następnej sekcji 1.4 omówione zostały pozostałe publikacje Autorki, a struktura rozprawy została przedstawiona w sekcji 1.5.

Rozdział 2. rozpoczyna się omówieniem tematyki oceny modeli (głębokiego) uczenia maszynowego – podkreślono najważniejsze wyzwania związane z wszechstronną ewaluacją takich metod. Autorka jasno przedstawia motywację i istotność podjętej tematyki badawczej oraz omawia wybrane metody XAI związane z głównymi zagadnieniami podjętymi w rozprawie. Rodział jest zwieńczony przedstawieniem „kontekstu” i wykorzystania informacji kontekstowej w analizie danych o różnych modalnościach, takich jak tekst, szeregi czasowe, sekwencje wideo czy dane obrazowe.

Kolejne **rozdziały 3–10** (z wyłączeniem rozdziału 6.) mają podobną strukturę – Doktorantka rozpoczyna każdy rozdział od krótkiego wprowadzenia, w którym omawia pytanie badawcze, które zostało podjęte w publikacji, odpowiadającej danemu rozdziałowi, a następnie przedstawia – często bardzo krótką – odpowiedź na postawione pytania. Taki zabieg był, moim zdaniem, bardzo trafny i ułatwia czytelnikowi przyswojenie treści. Po takim wprowadzeniu Autorka umieszcza treść (czasami rozszerzoną, tak jak w **rozdziale 4.**) wskazanej publikacji, której mgr inż. Paulina Tomaszewska jest współautorką.

W **rozdziale 3.**, Autorka przedstawia metodę LIMES – w opisywanych pracach, Doktorantka podjęła zagadnienia tworzenia modeli AI, które mają działać w systemach, w których dane stale napływają, a ich charakterystyka może się zmieniać (np. rozkład liczby przykładów z danych klas, w problemach klasyfikacyjnych – na takich skupiła się mgr inż. Paulina Tomaszewska). W pracy [1], Autorka opracowała lekkie techniki tworzenia metod AI dla napływających danych, jednocześnie wykorzystując kontekst czasowy takiego napływu. W zaproponowanej metodzie LIMES, modele są dynamicznie adaptowane, w zależności od rozkładu klas, który jest przewidywany na podstawie

danych historycznych. W rozdziale 3. Doktorantka krótko omawia inne podejścia stosowane w systemach, w których rozkład i charakterystyka danych zmienia się w czasie, trafnie uwypuklając ich zalety i wady, zwłaszcza związane z praktycznym wykorzystaniem technik uczenia ciągłego na urządzeniach brzegowych z ograniczeniami sprzętowymi. Badania eksperymentalne, w których Autorka porównuje metodę LIMES z technikami uczenia przyrostowego, komitetami klasyfikacyjnymi ze specjalizowanymi modelami oraz z modelami trenowanymi tylko na „najświeższych” danych (pomijając dane historyczne), skupiły się na klasyfikacji tweetów (tj. kraju ich pochodzenia). Wyniki wykazały, że LIMES może z powodzeniem konkurować z algorytmami znanymi z literatury, jednocześnie będąc metodą z potencjałem do wdrożenia na urządzeniach brzegowych.

W rozdziale 4., którego trzonem jest praca [2], Doktorantka omawia wpływ przestrzennej informacji kontekstowej w danych obrazowych na metody uczenia głębokiego oraz przedstawia metody XAI wykorzystywane do analizy modeli głębokich w analizie obrazów i podkreśla najważniejsze kierunki badań i luki badawcze, które powinny zostać zaadresowane w kontekście tworzenia metod XAI zorientowanych na analizę i zrozumienie zależności przestrzennych w danych obrazowych. Mgr inż. Paulina Tomaszewska zidentyfikowała zalety i wady istniejących metod oraz przekonująco omówiła kierunki dalszych badań, które powinny być realizowane, aby móc efektywnie analizować zależności przestrzenne, np. semantyczne i ich wpływ na jakość działania algorytmów uczenia głębokiego.

W kolejnym rozdziale 5., przedstawiona została metoda rozwiązywania testu matrycy Ravena (z ang. *Raven's Progressive Matrices*, RPM), która została opublikowana w pracy [3]. W opracowanej metodzie, w której wykorzystano informacje dotyczące kontekstu (o różnym zasięgu) i relacji przestrzennych pomiędzy obiektami w matrycy, Autorka połączyła głęboką ekstrakcję cech z takich matryc (z wykorzystaniem sieci głębokich typu *autoencoder*, w których cechy ekstrahowane są przez enkoder) z klasyfikatorem, którego celem jest stwierdzenie, które rozwiązanie testu jest lepsze (jeśli istnieje kilka rozwiązań, preferowane jest rozwiązanie „prostsze”). Po przeprowadzeniu porównań, w ramach których wybierany jest lepszy wariant rozwiązania (lub żaden, jeśli oba są niepoprawne), wyniki takich porównań są agregowane, a następnie wybierane jest najlepsze rozwiązanie. Autorka wszechstronnie przeanalizowała opracowaną technikę rozwiązywania testów matrycy Ravena i eksperymentalnie porównała ją z innymi algorytmami z literatury (z wykorzystaniem syntetycznie wygenerowanego zbioru danych). Wyniki eksperymentów wykazały, że wielostopniowy algorytm rozwiązywania RPM działa – w wielu wypadkach – lepiej niż inne algorytmy oparte na uczeniu głębokim. Uważam, że istotną zaletą opracowanego algorytmu *Duel-IQ* jest jego modułowość, która zresztą została eksperymentalnie przeanalizowana w badaniach eksperymentalnych.

Rozdział 6. jest krótkim wprowadzeniem do histopatologii cyfrowej, która będzie stanowić trzon kolejnych rozdziałów rozprawy, np. w kontekście analizy wysokorozdzielczych obrazów tkanki (z ang. *whole-slide image*, WSI). Istotnym elementem tego rozdziału jest wprowadzenie do metod uczenia wielokrotnej instancji (z ang. *multiple instance learning*, MIL), które są powszechnie wykorzystywane do analizy tego rodzaju danych. Autorka przedstawiła swoją obserwację dotyczącą sposobu agregacji cech z latek ekstrahowanych z WSI oraz tego, że w czasie takiej agregacji tracona jest informacja przestrzenna, w tym dotycząca zależności pomiędzy łatkami w WSI – taka informacja, według Doktorantki, może być przydatna w trakcie analizy WSI. Moim zdaniem, intuicja Autorki jest słuszna, również biorąc pod uwagę to, w jaki sposób analizowane są takie obrazy przez personel medyczny. W rozdziale 7., bazującym na pracy [4], Autorka przedstawia wyniki badań eksperymen-

talnych, w których analizowała wpływ kontekstu przestrzennego na jakość działania modeli uczenia głębokiego do klasyfikacji łatek obrazów histopatologicznych. W eksperymentach, w kontrolowany sposób usuwano elementy obrazu (wokół centralnej części łatki), które mogą potencjalnie zawierać informacje i cechy obrazowe istotne w czasie klasyfikacji. Wyniki wykazały, że ograniczenie kontekstu przestrzennego negatywnie wpłynęło na jakość działania analizowanych modeli.

Rozdział 8. opisuje opracowany przez Doktorantkę syntetyczny zbiór danych (*SyntheticSMIL*) zawierający wygenerowane dane o zróżnicowanych własnościach, wraz z konfigurowalną procedurą generowania danych syntetycznych oraz protokół *ReasonSMIL*, który może być wykorzystywany do oceny algorytmów MIL (bazujących na mechanizmach uwagi) [5]. Opracowane techniki zostały użyte do oceny dwóch algorytmów (CLAM i TransMIL), tj. do weryfikacji spójności ich predykcji i weryfikacji, czy predykcje wypracowane przez model bazują na „istotnych” łatkach wygenerowanego obrazu. Badania eksperymentalne wykazały, że modele różnią się pomiędzy sobą spójnością i sposobem wypracowywania predykcji. Uważam, że tworzenie takich generatorów danych syntetycznych o zadanej charakterystyce i jasno zdefiniowanych procedur eksperymentalnych są niezwykle ważne w kontekście analizy (jakościowej i ilościowej) modeli AI, które mogą być wszechstronnie zweryfikowane w kontrolowanym środowisku działania.

W **rozdziale 9.**, bazującym na pracy [6], Doktorantka omawia, w jaki sposób można byłoby wykorzystać zależności przestrzenne pomiędzy łatkami w obrazie WSI i analizuje, jak takie zależności, skwantyfikowane np. poprzez korelacje pomiędzy wektorami cech wyekstrahowanymi pomiędzy sąsiadującymi cechami, mogłyby zostać użyte w architekturach głębokich wzmocnionych takimi informacjami. Autorka wykazała, że te zależności przestrzenne są wystarczające do rekonstrukcji obrazu WSI.

Rozdział 10. opisuje metodę *DSCon* (*deep spatial context*), która została opublikowana w pracy [7]. Autorka zaproponowała wykorzystanie regresji do oceny istotności kontekstu przestrzennego w modelach MIL wykorzystujących mechanizmy uwagi – *DSCon* pozwala zweryfikować, czy (i w jakim stopniu) informacje przestrzenne są wykorzystywane przez model MIL (np. na poziomie etykiet, wyekstrahowanych cech czy błędów modelu). Mgr inż. Paulina Tomaszewska zaproponowała dodatkowo miary ilościowe, które kwantyfikują zachowanie i wykorzystanie kontekstu przestrzennego przez modele z uwagą na wspomnianych poziomach. Badania eksperymentalne skupiły się na wykorzystaniu metody *DSCon* do analizy algorytmów CLAM i TransMIL (w różnych wariantach). W ramach analizy wyników przeprowadzonych eksperymentów, Doktorantka zwróciła uwagę na ciekawe aspekty, np. na to, że globalny kontekst, w przypadku rozsianych zmian nowotworowych nie został zachowany (*DSCon* może pozwolić na zweryfikowanie tego, jakie relacje, np. pomiędzy łatkami w obrazie WSI zostały wyuczone przez model MIL). Warto podkreślić, że badania omówione w rozdziale 10. były bardzo wszechstronne i dogłębne.

W **rozdziale 11.**, Doktorantka skupia się na analizie wpływu wyboru zbioru treningowego (tj. łatek w wysokorozdzielczych obrazach WSI) na jakość działania modeli MIL (rozdział bazuje na artykule, który został zgłoszony na konferencję MICCAI 2026). Autorka skupiła się na analizie zmian typu *uveal melanoma* oraz zmian przerzutowych – detekcja takich przerzutów jest bardzo trudna, a predykcja ich pojawienia się, na podstawie charakterystyki zmiany pierwotnej, mogłaby istotnie ułatwić ten proces. W pracy zweryfikowano, czy modele MIL trenowane na losowych łatkach, na łatkach wybra-

nych przez lekarza (zgodnie z przyjętymi kryteriami) oraz na łatkach, którym model wykorzystujący mechanizm uwagi przypisał najwyższe jej wartości różnią się od siebie jakością działania. W tych eksperymentach Autorka również wykorzystwała modele CLAM i TransMIL. Eksperymenty wykazały, że jakość działania modelu CLAM jest zależna od zróżnicowania łatek treningowych, a zmniejszenie takiej różnorodności, np. poprzez ręczny wybór ROI, negatywnie wpływa na jakość działania modelu. Co ciekawe, łatki, którym przypisane zostały wysokie wartości uwagi nie były wybrane przez lekarzy w ramach obrysowanego ROI. Wyniki badań opisanych w tym rozdziale wykazały istotność doboru zbiorów treningowych w przypadku danych WSI i wskazują na to, że kosztowny proces ręcznego wyznaczania obszarów istotnych może negatywnie wpłynąć na jakość wyników modeli, poprzez zmniejszenie różnorodności zbioru treningowego.

Rozdział 12. jest podsumowaniem rozprawy, a w **rozdziale 13.** Autorka krótko omawia najistotniejsze kierunki dalszych badań.

Opracowane metody oraz otrzymane wyniki eksperymentalne przedstawione w rozprawie pozwalają mi z pełnym przekonaniem stwierdzić, że Doktorantka jest świetnie przygotowana do tego, żeby prowadzić dalsze badania związane z wyjaśnialną sztuczną inteligencją oraz projektowaniem, implementacją i oceną systemów sztucznej inteligencji w praktycznych zastosowaniach. Warto podkreślić, że Autorka udostępnia swoje implementacje i dba o reprodukowalność badań eksperymentalnych – jest to bardzo ważne w obecnym kryzysie reprodukowalności [8], z którym musimy się mierzyć.

Rozprawa doktorska mgr. inż. Pauliny Tomaszewskiej, a także artykuły, których jest współautorką prezentują dojrzałość badawczą oraz wiedzę teoretyczną i praktyczną Doktorantki w dyscyplinie *Informatyka Techniczna i Telekomunikacja*. Stronę merytoryczną rozprawy oceniam bardzo wysoko.

2.1 Pytania i uwagi

W trakcie lektury rozprawy nasunęły mi się następujące pytania i uwagi:

Rozdział 3.

1. Wyniki badań nad metodą LIMES (Table 3.2) pokazują, że wyniki dla LIMES i modelu trenowanego przyrostowo są bardzo zbliżone – czy z praktycznego punktu widzenia takie różnice są istotne? Czy Autorka rozważała przeprowadzenie testów statystycznych, które mogłyby zweryfikować, czy różnice pomiędzy tymi metodami są statystycznie istotne?
2. Na str. 48, Autorka opisuje procedurę walidacyjną i podziały danych, które zostały wykorzystane w badaniach eksperymentalnych. Moim zdaniem warto byłoby rozważyć wizualne przedstawienie takich podziałów i procedur walidacji analizowanych modeli – znacznie ułatwiłoby to zrozumienie treści.
3. W pracy korzystnie byłoby zamieścić przykłady poprawnie i niepoprawnie zaklasyfikowanych tweetów – czy, analizując takie przypadki, Autorka zauważyła jakieś interesujące zależności?
4. Autorka zauważyła, że w niektórych godzinach klasyfikacja tweetów (kraju ich pochodzenia) jest wyraźnie łatwiejsza dlatego, że bardzo dużo tweetów jest generowanych np. w USA lub w Brazylii. Czy Doktorantka analizowała pełne godzinowe rozkłady liczby przykładów z każdej klasy (kraju) oraz liczbę krajów, z których pochodzą tweety w kolejnych godzinach?

Rozdział 4.

1. Być może dobrym pomysłem byłoby podsumowanie, np. w formie tabeli, zbiorów danych omówionych w sekcji 4.5 (wraz z ich charakterystykami ilościowymi).
2. Umieszczenie przykładów wizualnych, np. podczas omawiania rozwiązań zorientowanych na analizę kontekstu w obserwacji Ziemi lub w obrazach histopatologicznych (str. 66) mogłoby pozwolić czytelnikowi na ich szybsze zrozumienie (i mogłoby pozwolić na jeszcze lepsze wypuklenie roli kontekstu i zależności przestrzennych w analizie tego rodzaju danych).

Rozdział 5.

1. Nie jest do końca jasne, w jaki sposób optymalizowana była architektura autoenkodera użytego do ekstrakcji cech z matryc Ravena (str. 83). Autorka wspomina o przeszukiwaniu siatki – warto byłoby podać dokładne hiperparametry (wraz z zakresami przeszukiwanych wartości), które były optymalizowane oraz jak dokładnie zdefiniowana była funkcja oceny wyników architektur. Czy, według Doktorantki, dałoby się ulepszyć proces doboru architektury/hiperparametrów autoenkodera, np. z wykorzystaniem metody typu AutoML/NAS? Jeśli tak, to w jaki sposób?

Rozdział 7.

1. Czy, według Doktorantki, obrazy, dla których ograniczono kontekst przestrzenny w eksperymentach (96×96) nie są zbyt małe, aby móc – na podstawie wyników eksperymentalnych, wyciągnąć wnioski, które można generalizować dla wysokorozdzielczych obrazów WSI?
2. Dla niektórych modeli, np. dla ResNeta 18 i dla DenseNeta 121 możemy zauważyć dość niestabilne zmiany metryk (np. precision, recall na Fig. 7.3; Autorka zauważa to zjawisko, i krótko wspomina o nim na str. 108). Czy Doktorantka próbowała przeanalizować, dlaczego tak się dzieje?
3. Korzystnie byłoby umieścić w pracy przykłady latek poprawnie i niepoprawnie zaklasyfikowanych w różnych konfiguracjach (ze zmienną wielkością zasłoniętego kontekstu). Czy takie analizy jakościowe/wizualne były prowadzone przez Doktorantkę? Czy udało zauważyć się tutaj jakieś ciekawe cechy takich obrazów (i czy takie przypadki były konsultowane z histopatologami)?

Rozdział 8.

1. W syntetycznie wygenerowanym benchmarku *SyntheticSMIL*, Autorka podzieliła zbiory, dla każdego z wariantów zbioru, na podzbiory *easy*, *medium* i *hard*, bazując na wartościach (wydaje się dość arbitralnych?) hiperparametrów użytych do generowania takich zbiorów. W jaki sposób zostały dobrane te wartości, i w jaki sposób Autorka oceniła trudność tych podzbiorów?

Powyższe uwagi i komentarze są raczej zachęceniem Doktorantki do rozmowy na temat opracowanych algorytmów i metod, i nie wpływają negatywnie na moją bardzo wysoką ocenę jakości rozprawy.

3 Ocena strony formalnej rozprawy doktorskiej

3.1 Ocena układu pracy

Układ rozprawy jest logiczny, a kolejne rozdziały wprowadzają czytelnika w najistotniejsze aspekty prac prowadzonych przez Doktorantkę.

3.2 Ocena zastosowanego piśmiennictwa

Bibliografia zawiera około 390 pozycji – jest świetnie skomponowana i jest aktualna, i nie pozostawia żadnych wątpliwości, że Doktorantka bardzo dobrze orientuje się w obecnym stanie wiedzy i najnowszych trendach związanych z tematyką rozprawy. Zauważyłem tylko bardzo drobne uchybienia, np. brak niektórych informacji bibliograficznych (np. brakujące strony w Abrahams et al., 2024; Adi et al., 2017; Chen et al., 2024; itp.).

3.3 Uwagi redakcyjne

Rozprawa jest napisana bardzo starannie, a tekst jest poprawny pod względem językowym, stylistycznym i interpunkcyjnym. Dopatrzyłem się kilku drobnych uchybień:

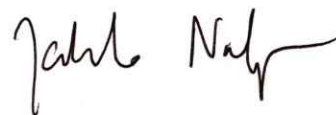
- Niektóre skróty są zdefiniowane wielokrotnie, np. LSTM na str. 28 i 33.
- Doktorantka miejscami używa brytyjskiego i amerykańskiego j. angielskiego (np. „*analysed*” i „*analyzing*” na str. 5; „e.g.” na str. 13 i „e.g.” na str. 23) – powinno to być uspołnione.
- Zauważyłem bardzo drobne niespójności w pisowni wielkimi/małymi literami, np. RED XAI i Red XAI (str. 5) i literówki, np. „*Such a model represent conditional...*” → „*Such a model represents conditional...*” (str. 44), „*Belowe*” (str. 65), „... *in the design of Duel-IQ*, ”.
- Pogrubienie najlepszych wyników w każdej z tabel ułatwiłoby ich szybsze przyswojenie (np. Tabela 3.2).
- Lepiej byłoby zapisywać np. 3×3 zamiast 3x3 (np. na str. 67).
- Wszystkie osie, na wszystkich rysunkach, powinny być opisane (np. Fig. 3.1, 3.2); powinny też być podane odpowiednie jednostki (np. [%] na Fig. 3.2).
- Nie znalazłem odniesienia z tekstu do Fig. 4.2.
- Zauważyłem, że „i.e.” jest kilkakrotnie niepoprawnie wykorzystane (np. na str. 73 czy 76).
- Na str. 119 Autorka wskazuje, że opracowany system jest zwizualizowany na rysunku Fig. 11.1 – powinno być Fig. 8.1 (na str. 134 czytamy, że „*For the analysis of TransMIL in Section 7...*” – chodzi o Section 8.7.2). Takie błędy wynikają z tego, że opisy zostały wzięte bezpośrednio z opublikowanych artykułów, ale nie wpływają negatywnie na łatwość przyswojenia treści rozprawy.

4 Konkluzja

Z pełnym przekonaniem stwierdzam, że przedmiotem rozprawy doktorskiej jest oryginalne rozwiązanie poprawnie zdefiniowanego problemu naukowego, a recenzowana dysertacja Pani mgr. inż. Pauliny Anny Tomaszewskiej **spełnia** wymagania stawiane rozprawom doktorskim przez Ustawę *Prawo o szkolnictwie wyższym i nauce*, Dz. U. 2023, poz. 742 z późn. zm. – praca prezentuje ogólną wiedzę teoretyczną i praktyczną Doktorantki w dyscyplinie *Informatyka Techniczna i Telekomunikacja* oraz umiejętność samodzielnego prowadzenia pracy naukowej.

W związku z powyższym, **wnioskuję o przyjęcie rozprawy doktorskiej oraz o dopuszczenie mgr. inż. Pauliny Tomaszewskiej do publicznej obrony.**

Ze względu na bardzo wysoki poziom prowadzonych badań i ich przekrojowość, oraz biorąc pod uwagę to, że prace, których Doktorantka jest współautorką, zostały opublikowane w prestiżowych czasopiśmie i konferencjach (m.in. ICML, position paper, IJCAI, doctoral consortium track, AISTATS, ICPR, czy Pattern Recognition), **wnioskuję o wyróżnienie rozprawy.**



Jakub Nalepa

Literatura

- [1] P. Tomaszewska, C. H. Lampert, Lightweight conditional model extrapolation for streaming data under class-prior shift, in: 2022 26th International Conference on Pattern Recognition (ICPR), 2022, pp. 2128–2134. doi:10.1109/ICPR56361.2022.9956195.
- [2] P. Tomaszewska, P. Biecek, Position: Do Not Explain Vision Models Without Context, in: R. Salakhutdinov, Z. Kolter, K. Heller, A. Weller, N. Oliver, J. Scarlett, F. Berkenkamp (Eds.), Proceedings of the 41st International Conference on Machine Learning, Vol. 235 of Proceedings of Machine Learning Research, PMLR, 2024, pp. 48390–48403.
URL <https://proceedings.mlr.press/v235/tomaszewska24a.html>
- [3] P. Tomaszewska, A. Żychowski, J. Mańdziuk, Duel-based Deep Learning system for solving IQ tests, in: G. Camps-Valls, F. J. R. Ruiz, I. Valera (Eds.), Proceedings of The 25th International Conference on Artificial Intelligence and Statistics, Vol. 151 of Proceedings of Machine Learning Research, PMLR, 2022, pp. 10483–10492.
URL <https://proceedings.mlr.press/v151/tomaszewska22a.html>
- [4] P. Tomaszewska, M. Sperkowski, P. Biecek, Does context matter in digital pathology? (2024). arXiv:2405.14301.
URL <https://arxiv.org/abs/2405.14301>

- [5] P. Tomaszewska, M. Gozdera, E. Sienkiewicz, P. Biecek, From Input to Insight: Probing the Reasoning of Attention-Based MIL Models, in: R. Guidotti, U. Schmid, L. Longo (Eds.), *Explainable Artificial Intelligence*, Springer Nature Switzerland, Cham, 2026, pp. 264–288.
- [6] P. Tomaszewska, Information injection to deep learning solutions in knowledge transfer, in: L. D. Raedt (Ed.), *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, International Joint Conferences on Artificial Intelligence Organization, 2022, pp. 5879–5880, doctoral Consortium. doi:10.24963/ijcai.2022/842.
URL <https://doi.org/10.24963/ijcai.2022/842>
- [7] P. Tomaszewska, E. Sienkiewicz, M. P. Hoang, P. Biecek, Deep Spatial Context: When Attention-Based Models Meet Spatial Regression (Digital Pathology), in: R. Bellazzi, J. M. Juarez Herrero, L. Sacchi, B. Zupan (Eds.), *Artificial Intelligence in Medicine*, Springer Nature Switzerland, Cham, 2025, pp. 438–447.
- [8] S. Kapoor, A. Narayanan, Leakage and the reproducibility crisis in machine-learning-based science, *Patterns* (Aug. 2023). doi:10.1016/j.patter.2023.100804.
URL [https://www.cell.com/patterns/abstract/S2666-3899\(23\)00159-9](https://www.cell.com/patterns/abstract/S2666-3899(23)00159-9)

