

Łódź, dn. 8 sierpnia 2025 r.

Prof. dr hab. inż. **Anna Fabijańska**
Politechnika Łódzka
Wydział Elektrotechniki, Elektroniki,
Informatyki i Automatyki
Instytut Informatyki Stosowanej
Ul. Stefanowskiego 18
90-537 Łódź



RECENZJA

wniosku w postępowaniu o nadanie

dr. inż. Maciejowi Piernikowi

stopnia doktora habilitowanego w dziedzinie nauk inżynieryjno-technicznych
w dyscyplinie informatyka techniczna i telekomunikacja

1. Podstawa sporządzenia recenzji

Podstawę opracowania niniejszej recenzji stanowiło pismo nr DIiT-64-01/2025 od Dziekana Wydziału Informatyki i Telekomunikacji Politechniki Poznańskiej, Pana dr hab. inż. Mikołaja Morzego, prof. Politechniki Poznańskiej z dnia 5 czerwca 2025 r., wystosowane w związku decyzją Rady Doskonałości Naukowej z dnia 15 maja 2025 r. oraz uchwałą Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Poznańskiej Nr 2025-07-055 z dnia 27 maja 2025 r., powołującymi mnie na recenzenta w postępowaniu o nadanie dr. inż. Maciejowi Piernikowi stopnia doktora habilitowanego w dziedzinie nauk inżynieryjno-technicznych, w dyscyplinie informatyka techniczna i telekomunikacja.

Recenzja sporządzona została na podstawie przekazanej mi w wersji elektronicznej dokumentacji przygotowanej przez Habilitanta, która obejmowała:

- wniosek przewodni,
- kopię dokumentu potwierdzającego posiadanie stopnia doktora,
- autoreferat,
- wykaz osiągnięć naukowych,
- listę publikacji wchodzących w skład osiągnięcia naukowego dołączonych do wniosku,
- kopie prac składających się na cykl publikacji stanowiących osiągnięcie naukowe,
- wykaz oraz oświadczenia współautorów publikacji będących w cyklu.

Podstawę prawną sporządzonej recenzji stanowiła ustawa z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce, w szczególności art. 219 ustawy.

2. Sylwetka habilitanta

Pan dr inż. Maciej Piernik jest absolwentem studiów inżynierskich (2009) oraz magisterskich (2010) na kierunku informatyka na Wydziale Informatyki Politechniki

Poznańskiej. Od roku 2008 jest zawodowo związany z tym wydziałem, gdzie przeszedł drogę od programisty zatrudnionego w projekcie do stanowiska adiunkta, na którym (począwszy od października 2015 roku) pracuje obecnie. Stopień doktora nauk technicznych w dyscyplinie informatyka Habilitant uzyskał na macierzystym wydziale w czerwcu 2015 roku, po przedłożeniu i obronie rozprawy nt.: „*Pattern-based clustering and classification of XML data*”.

W ogólnym ujęciu zainteresowania naukowe Habilitanta koncentrują się na rozwijaniu metod eksploracji danych i uczenia maszynowego, ze szczególnym uwzględnieniem ekstrakcji cech, klasyfikacji i analizy danych. Szczegółowa tematyka Jego badań dzieli się na dwa wyraźne etapy. Prace prowadzone przed uzyskaniem stopnia doktora koncentrowały się na przetwarzaniu danych strukturalnych w formacie XML, w szczególności na metodach klasyfikacji i klasteryzacji dokumentów XML oraz strumieni danych z dryfem koncepcji. Po uzyskaniu stopnia doktora Habilitant rozszerzył zakres badań na metody ekstrakcji cech i klasyfikacji w szerszym kontekście eksploracji danych, uwzględniając dane sekwencyjne, sieciowe, genetyczne i medyczne. Skupił się na projektowaniu algorytmów opartych na miarach odległości, statystyce rozkładów i metodach łączących różne reprezentacje danych, przy zachowaniu pewnej ciągłości wcześniejszych zainteresowań strukturami drzewiastymi.

Dorobek publikacyjny i wskaźniki bibliometryczne Habilitanta są raczej skromne biorąc pod uwagę etap kariery i okres działalności naukowej. Pan dr inż. Maciej Piernik jest autorem bądź współautorem 15 publikacji zaindeksowanych w bazie Scopus, w tym sześciu z okresu przed uzyskaniem stopnia doktora. Pośród Jego prac indeksowanych w bazie Scopus jest 8 artykułów w czasopismach posiadających współczynnik wpływu IF (*Lupus, Foundations of Computing and Decision Sciences, Bioinformatics, Molecular Diagnosis and Therapy, Knowledge and Information Systems, New Generation Computing, Knowledge Engineering Review*) oraz 7 referatów konferencyjnych. Zdecydowana większość tych prac wpisuje się w obszary tematyczne *computer science* oraz *mathematics*. Według bazy Scopus prace dra inż. Macieja Piernika były cytowane łącznie 95 razy, a wskaźnik Hirsha Habilitanta wynosi 6. Po wyłączeniu autocytowań, liczba cytowań i indeks H wynoszą odpowiednio 85 oraz 5. Prace Habilitanta nie spotkały się więc z istotnym zainteresowaniem środowiska naukowego.

Pan dr inż. Maciej Piernik uczestniczył w charakterze wykonawcy w pracach trzech zespołów badawczych realizujących projekty naukowe finansowane ze środków NCN (2) oraz MNiSW (1). Nie jest szerzej znany i zauważany w środowisku naukowym, o czym świadczy niewielka liczba recenzji dla czasopism, brak udziału w zespołach oceniających wnioski o finansowanie badań czy komitetach naukowych konferencji. Habilitant posiada natomiast doświadczenie we współpracy z sektorem gospodarczym. W szczególności pracował w firmie MNM Diagnostics nad rozwojem technologii wspomagającej diagnostykę i leczenie pacjentów nowotworowych gdzie przeszedł drogę od specjalisty ds. analizy danych do menedżera średniego i wysokiego szczebla. Aktualnie w ramach własnej działalności gospodarczej buduje i wdraża modele oparte o dane. Współuczestniczył również przy wdrożeniu kilku komercyjnych aplikacji internetowych wykorzystujących AI.

W ramach działalności dydaktycznej Pan dr inż. Maciej Piernik przygotował zajęcia z przedmiotu Internet Applications. Wypromował 8 inżynierów oraz 10 magistrów.

3. Stwierdzenie spełnienia przesłanki, o której mowa w art. 219 ust. 1, pkt. 1 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce dotyczącej posiadania stopnia doktora

Pan Maciej Piernik posiada stopień doktora nauk technicznych w dyscyplinie informatyka, który uzyskał 30 czerwca 2015 r. na Wydziale Informatyki Politechniki Poznańskiej. Kopia dyplomu stanowi załącznik do recenzowanego wniosku.

4. Ocena osiągnięcia naukowego wskazanego przez Habilitanta zgodnie z art. 219 ust. 1 pkt. 2 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce

4.1. Temat i publikacje wchodzące w skład osiągnięcia

Jako osiągnięcie naukowe wynikające z art. 219, pkt. 2 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce Pan dr inż. Maciej Piernik przedstawił cykl powiązanych tematycznie artykułów naukowych w czasopismach naukowych lub w recenzowanych materiałach z konferencji międzynarodowych nt.: *Metody klasyfikacji złożonych typów danych: podejścia oparte na cechach, odległościach i wzorcach*. Na cykl ten składa się siedem artykułów naukowych opublikowanych w latach 2017-2024, czyli w okresie po uzyskaniu przez Habilitanta stopnia doktora. Cykl publikacji obejmuje następujące artykuły naukowe:

1. **Maciej Piernik**, Tadeusz Morzy, Robert Susmaga, Izabela Szczęch: *On feature extraction using distances from reference points*, Foundations of Computing and Decision Sciences, 2024, vol. 49, no. 3, s. 287-302 (MNiSW₂₀₂₄: 40, IF: 1.3).
2. **Maciej Piernik**, Dariusz Brzeziński, Paweł Zawadzki: *Random Similarity Forests*, Proceedings of European Conference on Machine Learning and Knowledge Discovery in Database, ECML PKDD 2022, Part V, Springer, 2023, s. 53-69 (MNiSW₂₀₂₄: 140, CORE A).
3. **Maciej Piernik**, Dariusz Brzeziński, Paweł Sztromwasser, Klaudia Pacewicz, Weronika Majer-Burman, Michał Gniot, Dawid Sielski, Oleksii Bryzghalov, Alicja Woźna, Paweł Zawadzki: *DBFE: distribution-based feature extraction from structural variants in whole-genome data*, Bioinformatics, 2022, vol. 38, no. 19, s. 4466-4473 (MNiSW₂₀₂₄: 200, IF: 5.4).
4. **Maciej Piernik**, Tadeusz Morzy: *A study on using data clustering for feature extraction to improve the quality of classification*, Knowledge and Information Systems, 2021, vol. 63, no. 7, s. 1771-1805 (MNiSW₂₀₂₄: 100, IF: 3.1).
5. **Maciej Piernik**, Joanna Sołomiewicz, Arkadiusz Jachnik: *Assessing the Effectiveness of Sequences of Treatments Using Sequential Patterns*. Proceedings of 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Springer, 2019, s. 131-135 (MNiSW₂₀₂₄: 70, CORE B).
6. **Maciej Piernik**, Tadeusz Morzy: *Partial Tree-Edit Distance: A Solution to the Default Class Problem in Pattern-Based Tree Classification*, Proceedings of 21st Pacific-Asia

Conference on Advances in Knowledge Discovery and Data Mining, PAKDD 2017, Part II, Springer, 2017, s. 208-219 (MNiSW₂₀₂₄: 70, CORE B).

7. **Maciej Piernik**, Dariusz Brzeziński, Tadeusz Morzy, Mikołaj Morzy: *Using Network Analysis to Improve Nearest Neighbor Classification of Non-network Data*, Proceedings of 23rd International Symposium on Foundations of Intelligent Systems, ISMIS 2017, Springer, 2017, s. 105-115 (MNiSW₂₀₂₄: 20, CORE C).

4.2. Charakterystyka ilościowa osiągnięcia

Skład publikacji przedstawionych przez dr inż. Macieja Piernika do oceny w ramach powiązanego tematycznie cyklu kształtuje się więc następująco:

- 3 artykuły w czasopismach posiadających IF i uwzględnionych w bazie Journal Citation Reports (JCR);
- 4 referaty na konferencjach, w tym dwóch konferencjach uwzględnionych w bazie CORE Conference Ranking, jednej posiadającej ranking A, dwóch o rankingu B, oraz jednej o rankingu C.

Sumaryczny współczynnik *impact factor* (IF) publikacji wchodzących w skład osiągnięcia z uwzględnieniem aktualnych wartości wynosi 9.8, oraz 11.26 zgodnie z rokiem opublikowania. Sumaryczna liczba punktów MNiSW przypisanych do publikacji wchodzących w skład cyklu wynosi 640, przy czym punktacja poszczególnych prac waha się w zakresie od 20 do 200, osiągając średnią na poziomie 91 punktów oraz medianę na poziomie 70 punktów. **Od strony bibliometrycznej, ocena osiągnięcia naukowego przedstawionego przez dra inż. Macieja Piernika wypada więc słabo.**

Czasopisma, w których opublikowane zostały prace wchodzące w skład cyklu to prestiżowe czasopismo *Bioinformatics* (wyd. Oxford Academic) oraz dobrej renomy czasopisma *Knowledge and Information Systems* (wyd. Springer) i *Foundations of Computing and Decision Sciences* (wyd. Politechnika Poznańska). Międzynarodowe konferencje na których przedstawiono prace wchodzące w skład cyklu (*Conference on Machine Learning and Knowledge Discovery in Database*, *Conference on Artificial Intelligence in Medicine*, *Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, *International Symposium on Foundations of Intelligent Systems*) indeksowane są w bazie CORE Conference Ranking.

Pan dr inż. Maciej Piernik nie jest samodzielnym autorem żadnej z publikacji w cyklu, co jest dość nietypowe, gdyż zwyczajowo habilitanci zamieszczają takie publikacje w osiągnięciach przedkładanych do oceny. We wszystkich publikacjach w cyklu dr inż. Maciej Piernik jest pierwszym autorem, przy łącznej liczbie autorów od dwóch do dziesięciu (średnia: 4, mediana 3). Dla publikacji w cyklu nie określono procentowego udziału Habilitanta w poszczególnych publikacjach. Z załączonych do wniosku oświadczeń współautorów oraz zadeklarowanego wkładu merytorycznego Habilitanta wynika, że dr inż. Maciej Piernik jest pomysłodawcą lub opracował metody opisane w pięciu publikacjach. W publikacji [7] był jednym z trzech autorów deklarujących współautorstwo koncepcji/współudział w opracowaniu opisywanej metody, a w publikacji [4] – jednym z dwóch współautorów metody. W obu przypadkach, w dokumentacji nie wskazano, za które elementy metody odpowiadali

poszczególni współautorzy. Należy także odnotować, że do wniosku nie załączono oświadczeń wszystkich współautorów (w najlepszym bibliometrycznie artykule w *Bioinformatics* dot. metody BDFE [3] brakuje oświadczeń czterech współautorów). Habilitant był autorem bądź współautorem implementacji większości metod opisanych w publikacjach z cyklu oraz współuczestniczył w projektowaniu i przeprowadzaniu eksperymentów. **Indywidualny wkład merytoryczny dra inż. Macieja Piernika w większość prac wchodzących w skład cyklu zdaje się więc być wiodący, choć mógłby być on lepiej udokumentowany i doprecyzowany.**

Tematyka prac wchodzących w skład osiągnięcia naukowego przedstawionego do oceny w ogólnym ujęciu dotyczy metod ekstrakcji cech i klasyfikacji danych o złożonej strukturze, z wykorzystaniem miar podobieństwa, właściwości statystycznych rozkładów oraz analizy grafowej. Ten obszar tematyczny zaliczany jest do dyscyplin informatyka oraz informatyka techniczna i telekomunikacja. Sześć z siedmiu prac wchodzących w skład cyklu opublikowanych zostało w czasopismach lub materiałach konferencyjnych uwzględnionych w bazie bibliograficznej DBLP indeksującej publikacje z obszaru szerokorozumianej informatyki. Według bazy Scopus, wszystkie prace Habilitanta zamieszczone w cyklu należą do obszaru tematycznego *computer science*, natomiast pięć do obszaru tematycznego *mathematics*. Żadna z prac nie została zakwalifikowana do obszaru tematycznego *engineering*. **Osiągnięciu dra inż. Macieja Piernika bliżej jest więc do dyscypliny informatyka, chociaż można je również zakwalifikować do dyscypliny informatyka techniczna i telekomunikacja.**

4.3. Podsumowanie osiągnięcia

Tematyka przedstawionego przez dr inż. Macieja Piernika cyklu publikacji skupia się wokół różnorodnych metod analizy danych, przede wszystkim w kontekście eksploracji danych, ekstrakcji cech oraz technik poprawy skuteczności modeli uczenia maszynowego. Cykl jest **zasadniczo spójny**, choć obejmuje kilka różnych perspektyw. Osią wspólną jest **klasyfikacja danych z naciskiem na reprezentację i przetwarzanie danych złożonych lub o nietypowej reprezentacji**. W cyklu można wyróżnić trzy zasadnicze wątki:

- prace [1, 3, 4] dotyczą problemu ekstrakcji i transformacji cech,
- prace [2, 6, 7] proponują nowe metody klasyfikacji danych złożonych i mieszanych,
- praca [5] proponuje koncepcję analizy danych sekwencyjnych.

Ekstrakcja i transformacja cech

W pierwszym wątku, Habilitant koncentruje się na tworzeniu nowych reprezentacji danych poprzez ekstrakcję cech z oryginalnych danych, w sposób uniwersalny lub dostosowany do specyficznych typów danych.

Praca [1] proponuje metodę ekstrakcji cech z danych dowolnego typu opartą na odległości względem punktów referencyjnych. Metoda zakłada zbudowanie wektora cech dla każdego z punktów danych poprzez policzenie jego odległości od założonej liczby punktów referencyjnych. Wybór punktów referencyjnych realizowany na podstawie **centralności w grafie**, który reprezentuje relacje między wszystkimi obiektami w zbiorze danych. W szczególności, wierzchołkami w grafie są obiekty zbioru danych, które połączone z wszystkimi innymi wierzchołkami tworzą graf pełny. Krawędzie opisane są wagami

uwzględniającymi zarówno podobieństwo obiektów jak i przynależność klasową. Punkty referencyjne wybierane są w oparciu o centralność wierzchołków. Habilitant zauważa, że wybór top-k punktów o największej centralności może prowadzić do nadmiarowości z uwagi na bliskie sąsiedztwo punktów w przestrzeni cech. Dlatego, proponuje podejście **θ -neighborhood** pozwalające znaleźć punkty dominujące lokalnie. W szczególności, dla danego punktu sprawdzane są pozostałe punkty i odrzucane są te, które mają inny punkt bliżej siebie w tym samym ogólnym kierunku. Kierunek definiuje się geometrycznie jako obszar w kształcie rozwartych trójkątów z kątem θ u podstawy. Jeśli znajdzie się punkt, który leży bliżej i w takim samym kierunku, to punkt nie jest lokalnym maksimum. Jeśli nie - uznaje się go za dominujący lokalnie i wybiera jako punkt referencyjny. Dzięki temu liczba punktów referencyjnych jest dobierana automatycznie. Autorskie podejście do wyboru punktów referencyjnych Habilitant porównał z wyborem losowym oraz wykorzystaniem centrów klastrów. Testował również różne miary centralności. Eksperymenty na 15 zbiorach danych wykazały, że proponowana metoda ekstrakcji cech jest uniwersalna. Nowy sposób wyboru punktów referencyjnych daje lepsze rezultaty niż klasyczny wybór punktów o najwyższych wartościach miary centralności. W praktyce liczba punktów referencyjnych ma większe znaczenie niż sposób ich wyboru, jeżeli tylko są odpowiednio rozproszone w przestrzeni danych. W pracy Habilitant proponuje również neuron euklidesowy tj. nowy typ neuronu którego wyjściem nie jest klasyczna suma ważona cech z aktywacją, ale odległość Euklidesowa pomiędzy wektorem wejściowym a wektorem wag neuronu. Brakuje eksperymentów, które pokazałyby rzeczywistą skuteczność neuronu euklidesowego lub zastosowanie w sieciach głębokich. **Pomysł pozostaje więc jedynie na poziomie koncepcji.**

W pracy [3] Habilitant proponuje zestaw trzech metod ekstrakcji cech z danych genetycznych pod nazwą DBFE (ang. *distribution-based feature extraction*) opartych na rozkładzie długości wariantów strukturalnych genomu. Pierwszy wariant, tj. podział względem kwantyli (ang. *quantile binning*) dzieli dane tak, by w każdym przedziale była podobna liczba wariantów. Przedziały są ustalane na podstawie rozkładu długości w całym zbiorze, bez uwzględniania etykiet klas. Metoda działa w sposób nienadzorowany. Drugi wariant, również nienadzorowany, bazuje na założeniu, że długości wariantów mają naturalne skupiska. Dlatego wyznacza przedziały na podstawie mieszaniny rozkładów Gaussa dopasowanej do rozkładu długości wariantów z wykorzystaniem algorytmu EM. Wynikowe klastry odpowiadają przedziałom, których licznosc używana jest jako cechy. Trzecie podejście – KDE (ang. *Kernel Density Estimation*) dokonuje ekstrakcji cech w sposób nadzorowany wykorzystując długości wariantów strukturalnych do oszacowania rozkładów gęstości osobno dla każdej klasy. Granice przedziałów definiuje się jako punkty przecięcia tych rozkładów. Każda cecha opisuje liczbę lub udział wariantów długości mieszczących się w danym przedziale. Dzięki temu cechy mają wysoką moc dyskryminacyjną, ponieważ są skonstruowane tak, by najlepiej odróżniać klasy. Metoda automatycznie ogranicza liczbę cech i zapewnia ich interpretowalność. Weryfikacja eksperymentalna skuteczności wariantów DBFE jako ekstraktorów cech została przeprowadzona na trzech rzeczywistych zbiorach danych nowotworowych oraz z uwzględnieniem kilku popularnych klasyfikatorów. Dokonano również porównania z cechami ekspercko tworzonymi oraz z klasycznymi biomarkerami PAM50. Wyniki pokazały, że DBFE umożliwia automatyczne generowanie cech o dobrej wartości predykcyjnej, często

przewyższającej klasyczne biomarkery. Najlepsze wyniki osiągnięto jednak przy połączeniu cech DBFE z biomarkerami. **Proponowanej przez Habilitanta metody nie odniesiono jednak do innych metod ekstrakcji cech z genomów.**

W pracy [4] Habilitant testuje w szczególności znaną z literatury koncepcję wykorzystania wyników klasteryzacji do poprawy wyników klasyfikacji. W szczególności, rozważa możliwość przekształcenia wyników klasteryzacji w nowe cechy wspomagające klasyfikację. W szczególności, zaproponował rozwiązanie, w którym nowe cechy uzyskiwane są na podstawie przynależności danych do klastrów lub odległości do ich reprezentantów (np. centroidów). Cechy te następnie wykorzystał podczas trenowania klasyfikatorów samodzielnie lub rozszerzając oryginalny zestaw cech. Walidację rozwiązania przeprowadził w wielu wariantach na 16 publicznie dostępnych zbiorach danych. Sprawdził skuteczność wykorzystania m.in. różnych reprezentacji przynależności do klastrów, algorytmów klasteryzacji, sposobów tworzenia klastrów oraz różnych metryk. Zbadał także skuteczność zestawów cech oryginalnego, nowego i rozszerzonego w eksperymentach z siedmioma różnymi algorytmami klasteryzacji. Dzięki temu ustalił najlepszą konfigurację metody. Eksperymenty wykazały, że samodzielne wykorzystanie uzyskanych przez proponowane podejście cech nie poprawia dokładności klasyfikacji. Cechy z klasteryzacji są przydatne jako uzupełnienie oryginalnych, a nie jako ich zamiennik. Jednak efektywność wykorzystania rozszerzonego zestawu cech zależy od klasyfikatora, zbioru danych i liczby klastrów. W szczególności, technika ta może poprawiać skuteczność klasyfikatorów, ale tylko w wybranych scenariuszach – głównie z liniowymi modelami i przy odpowiedniej liczbie klastrów.

Metody klasyfikacji danych złożonych i mieszanych

W drugim wątku Habilitant proponuje nowe algorytmy klasyfikacji dedykowane do przetwarzania danych o złożonej reprezentacji opisanych cechami o różnym typie i skali.

W pracy [2] Habilitant proponuje nową metodę klasyfikacji RSF (ang. *Random Similarity Forest*) strukturę drzew decyzyjnych przeznaczoną do danych o mieszanych typach cech; zarówno skalarnych, jak i złożonych (grafy, rozkłady, sekwencje). RSF łączy zalety znanych metod Random Forests (RF) i Similarity Forests (SF) radząc sobie z cechami różniącymi się typem, skalą i miarą odległości. Kluczową nowością jest podział węzłów drzewa oparty na projekcjach tworzonych indywidualnie dla każdej cechy, z wykorzystaniem dedykowanych miar. W przeciwieństwie do SF, które traktują całe przykłady jako obiekty i obliczają jedną odległość globalną, RSF analizuje każdą cechę indywidualnie z dedykowaną miarą odległości, nawet jeśli reprezentuje obiekt złożony (np. sekwencję lub graf). Budowanie drzewa rozpoczyna się od losowego wyboru cech, a następnie dla każdej z nich losowane są przykłady z różnych klas, które służą jako punkty referencyjne do tworzenia jednowymiarowej projekcji wszystkich pozostałych przykładów w węźle. Różnice odległości do tych punktów porządkują dane, umożliwiając podział minimalizujący indeks Giniego. Proces powtarza się w kolejnych węzłach aż do uzyskania stabilnych liści. Metodę zweryfikowano na 30 zbiorach danych: skalarnych, złożonych (sekwencje, grafy, szeregi czasowe) i mieszanych. RSF osiąga lepsze wyniki niż RF i SF na danych złożonych i mieszanych; na danych czysto skalarnych jest porównywalny z RF, choć czasem nieco gorszy. Szczególnie dobrze sprawdza się w danych

multi-omics, gdzie różne typy cech wymagają różnych miar. Ograniczeniami RSF są: konieczność doboru miar odległości dla każdej cechy, koszt obliczeniowy zależny od kosztu wyznaczenia odległości, wysoka losowość (wymagająca wielu drzew), ograniczenie do klasyfikacji binarnej oraz brak mechanizmu ważenia cech i analizy skalowalności.

Praca [6] dotyczy problemu klasyfikacji struktur drzewiastych (reprezentowanych w pracy przez dokumenty XML) na podstawie poddrzew. W szczególności, Habilitant wprowadza nową miarę PTED (ang. *partial tree-edit distance*) odległości między drzewami, która ocenia minimalny koszt transformacji drzewa wzorca w dowolne poddrzewo danych, uwzględniając usuwanie liści i zmianę etykiet. Habilitant proponuje również algorytm obliczania PTED oparty na programowaniu dynamicznym. Eksperymentalna weryfikacja na danych syntetycznych i rzeczywistych wykazała, że zastosowanie PTED poprawia dokładność klasyfikacji w porównaniu z tradycyjnymi podejściami opartymi na klasie domyślnej. Pomimo dobrych wyników, złożoność obliczeniowa $O(n^4)$ może ograniczać zastosowanie PTED do dużych drzew. **Brak porównania z osadzeniami drzew, czy metodami wykorzystującymi sieci neuronowe do uzyskania osadzeń jest zauważalnym brakiem.**

W pracy [7] Habilitant przedstawia podejście do poprawy skuteczności klasyfikatora k-najbliższych sąsiadów. W szczególności, proponuje wprowadzenie wag dla przykładów uczących. Wagi te nie bazują na klasycznych miarach odległości, lecz wykorzystują analizę sieciową i miary centralności w grafach. Habilitant współ-opracował nową metodę ważenia instancji opierającą się na strukturze grafu utworzonego z danych, które pierwotnie nie stanowiły sieci. W szczególnym ujęciu budowany jest graf k-najbliższych sąsiadów, w którym każdy wierzchołek reprezentuje jedną instancję treningową. Krawędzie grafu ważone są na podstawie podobieństwa między instancjami i dodatkowo oznaczone znakiem dodatnim (dla instancji tej samej klasy) lub ujemnym (dla instancji różnych klas). Następnie dla wszystkich wierzchołków wyznaczane są miary centralności (tj. ich istotności w grafie), które w klasyfikatorze k-NN pełnią rolę wag instancji treningowych. Weryfikację eksperymentalną metody przeprowadzono na 30 popularnych zbiorach danych z repozytorium UCI, obejmujących problemy klasyfikacji binarnej i wieloklasowej, zróżnicowanych pod względem liczby instancji, cech i klas. W analizie centralności Habilitant uwzględnił cztery miary, stopień (*degree*), współczynnik klasteryzacji (*clustering coefficient*), status oraz centralność pozytywno-negatywną (PN). **Zaproponowana metoda nie wykazała jednoznacznej przewagi nad nieważonym k-NN we wszystkich rozważanych przypadkach.** Niemniej jednak w wybranych scenariuszach poprawiała jakość klasyfikacji, zwłaszcza, gdy zastosowano miarę centralności PN.

Koncepcja analizy danych sekwencyjnych

W trzecim wątku badawczym mieści się praca [5], w której Habilitant wprowadza koncepcję wzorców sekwencyjnych zmieniających stan (ang. *state-changing sequential patterns*) w celu oceny skuteczności sekwencji zabiegów medycznych w kontekście zmian stanu zdrowia pacjenta mierzonego określonymi wskaźnikami. Metoda łączy dwa strumienie danych, tj. wydarzenia medyczne i stany pacjenta po nich występujące. Następnie dzieli je na podsekwencje według kierunku zmian stanu zdrowia, tworząc dwa zbiory *poprawa*

i pogorszenie. W obu zbiorach identyfikuje te wzorce, które występują znacznie częściej każdym ze zbiorów wskazując odpowiednio na poprawę lub pogorszenie stanu pacjenta. **Praca ma charakter koncepcyjny i nie przedstawia dojrzałego rozwiązania algorytmicznego. Główną wartością pracy jest zidentyfikowanie problemu i zaproponowanie wstępnej koncepcji jego rozwiązania. Ponadto, eksperyment proof-of-concept przeprowadzony został na jednym zbiorze danych w sposób niewystarczający do wykazania skuteczności metody.**

Podsumowując, osiągnięcia dra inż. Macieja Piernika przedstawione w ocenianym cyklu publikacji obejmują:

- Opracowanie metody ekstrakcji cech z dowolnego typu danych, bazującą na odległości od punktów referencyjnych wybranych z użyciem centralności w grafie [1].
- Opracowanie algorytmu RSF (ang. *Random Similarity Forests*) pozwalającego na klasyfikację danych z mieszanymi typami cech [2].
- Opracowanie trzech metod DBFE (ang. *Distribution-Based Feature Extraction*) do automatycznego generowania cech z rozkładów długości wariantów strukturalnych genomu [3].
- Zbadanie i wykazanie, czy i kiedy wyniki klasteryzacji mogą poprawiać klasyfikację, służąc jako źródło dodatkowych cech [4].
- Wprowadzenie koncepcji wzorców sekwencyjnych zmieniających stan (ang. *State-Changing Sequential Patterns*) do oceny skuteczności sekwencji zabiegów medycznych [5].
- Opracowanie miary PTED (ang. *Partial Tree-Edit Distance*) oraz algorytmu jej obliczania z przeznaczeniem do klasyfikacji struktur drzewiastych [6].
- Propozycję metody ważenia instancji w klasyfikatorze k-najbliższych sąsiadów wykorzystującej analizę centralności w grafach [7].

Poszczególne autorskie rozwiązania Habilitanta różnią się poziomem dojrzałości. Metody opisane w pracach [1], [2] (RSF), [3] (DBFE) czy [7] charakteryzują się największą kompletnością, zawierając formalizację metod, ich analizę oraz weryfikację eksperymentalną na popularnych zbiorach danych. Nie zawsze jednak odnoszą się do metod konkurencyjnych, innych niż bazowe. Średni poziom dojrzałości charakteryzuje rozwiązanie PTED [6], które posiada ograniczoną weryfikację eksperymentalną, w szczególności przetestowane zostało tylko w jednym scenariuszu (klasyfikacja dokumentów XML) oraz nie zostało porównane z alternatywnymi rozwiązaniami pozyskiwania reprezentacji, np. osadzania grafów/drzew. Najmniej dojrzałą pracą jest [5], która opisuje metodę w wczesnym stadium rozwoju, na poziomie pojedynczego eksperymentu proof-of-concept, z brakiem kompletnego opisu metody oraz przekonującej eksperymentalnej weryfikacji. Podobnie idea neuronu euklidesowego proponowana w pracy [1] pozostaje na poziomie koncepcji, która nie została przetestowana.

Praca [4] zasadniczo nie proponuje nowych rozwiązań, a jedynie systematycznie sprawdza i weryfikuje pomysł z literatury, który wcześniej nie był kompleksowo zbadany.

Z uwagi na charakter konferencyjny, prace [2],[5],[6],[7] pobieżnie odnoszą się do aktualnego stanu wiedzy i konkurencyjnych rozwiązań, nie przedstawiając szerszego kontekstu. Autoreferat Habilitanta (który zresztą jest dość pobieżny) nie uzupełnia tej luki. Ponadto, w cyklu publikacji Habilitant koncentruje swoją uwagę wyłącznie na klasycznych metodach uczenia maszynowego, co zasadniczo ma uzasadnienie w mniejszej ilości danych treningowych i większej wytłumaczalności modeli niż w przypadku modeli sieci neuronowych. Habilitant zdaje się jednak realizować swoje prace bez świadomości szerszego kontekstu związanego z rozwojem metod wykorzystujących głębokie sieci neuronowe. Natomiast niektóre z prac Habilitanta mogłoby zyskać na porównaniu ze współczesnymi modelami uczenia głębokiego, które również adresują problemy związane z danymi złożonymi i ekstrakcją cech. Przykładowo metoda DBFE nie została odniesiona do metod bazujących na uczeniu głębokim popularnych w analizie danych genomicznych. Analogicznie, klasyczne podejścia bazujące na edycji drzew [6] czy wykorzystujące grafową/sieciową reprezentację [1],[7] nie zostały odniesione do współczesnych podejść opartych na grafowych sieciach neuronowych lub sieci typu *tree-structured neural network*. Takie porównanie zdecydowanie lepiej osadziłoby prace Habilitanta w aktualnym stanie wiedzy oraz pozwoliłyby na ocenę wartości dodanej proponowanych metod w kontekście zarówno klasycznych, jak i najnowszych podejść w dziedzinie.

Należy również odnotować, że część prac Habilitanta nie dostarcza jednoznacznych dowodów na przewagę proponowanych rozwiązań nawet nad klasycznymi metodami. W pracy [1] różne strategie wyboru punktów referencyjnych nie dały statystycznie istotnej przewagi w porównaniu do losowego wyboru, a kluczowe okazało się jedynie odpowiednie dobranie liczby tych punktów. Praca [4] nie identyfikuje jednej najlepszej konfiguracji podejścia wykorzystującego klasteryzację do ekstrakcji cech, lecz zestaw wniosków zależnych od kontekstu. Eksperymenty wykazały jedynie subtelne różnice w wynikach, a w niektórych przypadkach proponowane cechy nawet pogarszały jakość klasyfikacji dla konkretnych modeli. Również w pracy [7] proponowana metoda ważenia danych w klasyfikatorze k-NN nie wykazała jednoznacznej przewagi nad wersją bazową.

Z uwagi na brak jednoznacznej demonstracji przewagi proponowanych rozwiązań nad istotnymi rozwiązaniami konkurencyjnymi, niewystarczające odniesienie do najnowszych osiągnięć w dziedzinie oraz zróżnicowany poziom dojrzałości proponowanych rozwiązań, należy stwierdzić, że **dr inż. Maciej Piernik nie wykazał w sposób przekonujący, że Jego osiągnięcie naukowe w istotnym stopniu poszerza aktualny stan wiedzy w dyscyplinie informatyka techniczna i telekomunikacja.**

5. Ocena aktywności naukowej Habilitanta zrealizowanej w więcej niż jednej uczelni, instytucji naukowej lub instytucji kultury, w szczególności zagranicznej zgodnie z art. 219 ust. 1 pkt. 3 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce

Dr inż. Maciej Piernik nie wykazał w sposób przekonujący swojej aktywności naukowej zrealizowanej w więcej niż jednej uczelni, instytucji naukowej lub instytucji kultury. Jako aktywności w tym obszarze Habilitant wskazuje swoją działalność recenzencką, współ-opiekę nad aplikacjami i bazami danych wchodzącymi w skład Polskiego Rejestru Wrodzonych Wad

Rozwojowych (PRWWR) oraz publikacje przygotowane we współpracy z naukowcami z różnych instytucji medycznych.

Aktywność recenzencką Habilitanta dla czasopism międzynarodowych i konferencji trudno uznać za aktywność naukową zrealizowaną w uczelni lub instytucji naukowej. Jest to działalność wspierająca system naukowy, ale nie twórcza działalność naukowa sama w sobie, gdyż nie wiąże się z wytworzeniem własnego dorobku naukowego. Wydawcy natomiast to podmioty komercyjne lub organizacje zawodowe, niebędące jednak instytucjami naukowymi.

Z kolei opis zaangażowania dra inż. Macieja Piernika w działanie Polskiego Rejestru Wrodzonych Wad Rozwojowych wskazuje na charakter inżynierski oraz administratorski jego prac, niespełniający warunków aktywności naukowej. Habilitant nie dostarcza również potwierdzenia tej aktywności, natomiast strona internetowa rejestru nie odnotowuje Jego udziału.

Współpracę dra inż. Macieja Piernika z naukowcami z instytucji medycznych również trudno uznać za aktywność naukową realizowaną w więcej niż jednej uczelni lub instytucji naukowej. W publikacjach będących wynikiem tej współpracy Habilitant afiliuje do swojej macierzystej jednostki. Do wniosku nie dołączył również żadnych dokumentów wskazujących na to, że współpracę realizował formalnie działając badawczo w dwóch różnych instytucjach naukowych (np. w ramach projektu, zatrudnienia, czy stażu naukowego).

Ww. aktywności są w mojej ocenie niewystarczające, aby uznać, że dr inż. Maciej Piernik wykazuje się istotną aktywnością naukową realizowaną w więcej niż jednej uczelni, instytucji naukowej lub instytucji kultury.

6. Podsumowanie

W podsumowaniu niniejszej recenzji stwierdzam, że zarówno osiągnięcie naukowe dra inż. Macieja Piernika w postaci cyklu powiązanych tematycznie artykułów naukowych, jak i Jego aktywność naukowa zrealizowana w więcej niż jednej uczelni, instytucji naukowej lub instytucji kultury, w szczególności zagranicznej **nie spełniają wymagań** określonych w art. 219 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym, stawianych kandydatom ubiegającym się o nadanie stopnia naukowego doktora habilitowanego, w tym wypadku nauk inżynieryjno-technicznych w dyscyplinie informatyka techniczna i telekomunikacja.

W związku z powyższym, wniosek dra inż. Macieja Piernika o nadanie stopnia doktora habilitowanego uważam za przedwczesny i opiniuję negatywnie.

Anne Falejczyk