

Opole, 30.12.2023 r.

Recenzja pracy doktorskiej (sporządzonej w oparciu o cykl publikacji) mgr inż. Anny Czmił pt.: „Usprawnienie procesu diagnostyki medycznej przy użyciu metod sztucznej inteligencji” przygotowanej pod kierunkiem promotora: dr hab. Inż. Damiana Mazura, prof. PRz oraz Kopromotora: Dr hab. n. med. Bogdana Obrzuta, prof. UR, w dyscyplinie: Informatyka Techniczna i Telekomunikacja

Recenzja została sporządzona w związku z powołaniem przez Radę Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Rzeszowskiej im. Ignacego Łukasiewicza w dniu 11.10.2023 roku do pełnienia funkcji recenzenta w postępowaniu o nadanie stopnia naukowego doktora nauk technicznych pani mgr inż. Annie Czmił.

Niniejsza recenzja ma za zadanie zgodnie z Art. 13 ust. 1 ustawy z dnia 14 marca 2003 r. o stopniach i tytule naukowym oraz o stopniach i tytule naukowym w zakresie sztuki (t.j. Dz.U. 2017 poz. 1789, dalej jako: u.s.n.), mającego zastosowanie w sprawie w związku z art. 175 ust. 1 Przepisów wprowadzających ustawę – Prawo o szkolnictwie wyższym i nauce z 3.7.2018 r. (Dz.U. 2018 r. poz. 1669) ocenić czy rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego oraz wkładu w dyscyplinę.

W ramach przeprowadzonej recenzji zostaną ocenione następujące punkty:

1. Tematyka pracy doktorskiej i jej wkład w dyscyplinę.

Ad. 1. Temat pracy doktorskiej brzmi: „Usprawnienie procesu diagnostyki medycznej przy użyciu metod sztucznej inteligencji”.

Rozprawa ta składa się z **pięciu** (opublikowanych w latach 2019-2023) powiązanych tematycznie artykułów naukowych w języku angielskim i jest zgodna z przepisami zawartymi w artykule 187 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce. Artykuły stanowiące niniejszy cykl publikacji zostały już wcześniej poddane recenzji w procesie publikacyjnym i prezentują nowoczesne rozwiązania związane z tematyką pracy doktorskiej, która to koncentruje się na wspomaganie i doskonaleniu procesu diagnostyki medycznej w oparciu o zastosowanie metod sztucznej inteligencji.

W mojej ocenie – wybrany temat rozprawy, jest aktualny i stanowi istotny wkład dla dyscypliny naukowej: **Informatyka Techniczna i Telekomunikacja**.

2. Zagadnienia naukowe rozprawy – cel i teza pracy.

WPEŁYNIŁO

18. STY. 2024



Ad. 2. W pracy została postawiona następująca hipoteza:

- *Możliwe jest wykorzystanie różnych metod sztucznej inteligencji do analizy danych medycznych i automatyzacji wybranych procesów diagnostycznych, pozwalające na uzyskanie interpretowalnych wyników z dokładnością i efektywnością nie gorszą niż innych istniejących metod znanych z literatury.*

W celu potwierdzenia hipotezy postawionej przez autorkę sformułowano **5** zadań szczegółowych:

- I. *Zastosowanie metod sztucznej inteligencji do klasyfikacji cukrzycy typu 1 na podstawie danych uzyskanych za pomocą nieinwazyjnych pomiarów aktywności fizycznej.*
- II. *Opracowanie metody pozwalającej na automatyczne, jednoczesne rozpoznawania i zliczanie czerwonych i białych krwinek oraz płytek krwi na podstawie zdjęć mikroskopowych z wykorzystaniem głębokich sieci neuronowych.*
- III. *Opracowanie aplikacji umożliwiającej automatyzację procesu oceny, składania i identyfikacji sekwencji genomowych uzyskanych za pomocą nowych metod sekwencjonowania przy użyciu narzędzi korzystających z metod uczenia maszynowego.*
- IV. *Implementacja w języku Python klasyfikatora opartego na logice rozmytej i programowaniu ekspresji genów, służącego do generowania wysoce interpretowalnych reguł rozmytych.*
- V. *Opracowanie narzędzia pozwalającego na eksperymentalne porównanie wybranych rozmytych algorytmów opartych na regułach do klasyfikacji danych medycznych.*

Każdy wskazany przez panią mgr inż. Annę Czmił cel odpowiada zadaniu, które zostało zrealizowane i opublikowane w kolejnych artykułach naukowych z serii publikacji.

Na podstawie wyników przedstawionych w artykułach, stanowiących jednotematyczny cel publikacji, uważam, że cele/zadania zostały potwierdzone a hipoteza spełniona. Oceniając otrzymane przed doktorantką wyniki, można stwierdzić, że niniejsza rozprawa doktorska spełnia wszystkie standardy obowiązujące w przypadku prac doktorskich oraz, iż w znacznym stopniu przyczynia się do rozwoju dyscypliny naukowej jaką jest **Informatyka Techniczna i Telekomunikacja**.

3. Struktura pracy

Ad. 3. W ramach realizacji niniejszej pracy doktorskiej podjęto się tematyki związanej z wykorzystaniem metod sztucznej inteligencji w celu wspomaganie i doskonalenie procesu diagnostyki medycznej. Praca składa się ze **163** stron (wliczając stronę tytułową, podziękowania, oświadczenia współautorów oraz artykuły stanowiące cykl), podzielona została na **10** rozdziałów (w tym nieponumerowana literatura, dorobek naukowy autorki, wykaz stosowanych oznaczeń, artykuły wchodzące w skład cyklu, streszczenie w języku polskim, streszczenie w języku angielskim, oświadczenia współautorów).

Jako, że praca ta składa się głównie z cyklu publikacyjnego, ma ona nieco inną strukturę niż tradycyjne dysertacje. Pierwsza część to wprowadzenie teoretyczne i jego struktura zbliżona jest do tradycyjnych rozpraw. Składa się ona z **4** rozdziałów i opisana jest na **56**

stronach. Literatura zawiera **87** pozycji literaturowych, **5** rysunków, **3** tabel, **1** listingu i **1** wzoru.

Pierwszą część pracy rozpoczyna "**Wprowadzenie**" - rozdział **pierwszy**, w którym zawarto w nim wstęp do tematu, skład publikacji wchodzących do cyklu, motywację oraz aktualny stan wiedzy, hipotezę badawczą oraz cele pracy (zadania szczegółowe). Rozdział **drugi** opisuje proponowane zastosowania metody opartej o sztuczną inteligencję i ich wykorzystanie w diagnostyce medycznej. Rozdział ten składa się z **5** podrozdziałów, gdzie w **pierwszym** z nich opisano nieinwazyjną metodę wykrywania cukrzycy typu 1; w **drugim** przedstawiono problematykę rozpoznawania krwi na podstawie zdjęć z jej rozmazu z wykorzystaniem głębokiej sieci neuronowej; **trzeci** podrozdział przedstawia możliwości automatyzacji procesów oceny jakości genomów prokariotycznych; w **czwartym** podrozdziale przedstawiono zastosowania programowania ekspresji genów w celu wydobycia metareguł z danych medycznych; a w **piątym**, ostatnim, porównano rozmaite klasyfikatory oparte zarówno na regułach i metaregułach. Rozdział **trzeci** zawiera podsumowanie i wnioski, gdzie opisane realizację poszczególnych zadań, wkład autorki w cykl publikacji oraz kierunki dalszych badań. Rozdział **czwarty** (nieponumerowany) stanowi literatura, na którą składa się **87** pozycji z czego **57** stanowią najnowsze publikacje z lat 2017-2023.

Druga część pracy - cykl publikacyjny składa się z pięciu artykułów naukowych opublikowanych w renomowanych czasopismach z wysokimi współczynnikami Impact Factor (IF):

- I. **Czmił, A.** Czmił, S., & Mazur, D. (2019). *A Method to Detect Type 1 Diabetes Based on Physical Activity Measurements Using a Mobile Device*. *Applied Sciences*, 9(12), 2555. doi:10.3390/app9122555. IF (2019): 2,474, liczba punktów: 70, wkład: 33,33%.

Tytuł zadania: Zastosowanie metod sztucznej inteligencji do klasyfikacji cukrzycy typu 1 na podstawie danych uzyskanych za pomocą nieinwazyjnych pomiarów aktywności fizycznej.

Praca dotyczy nowatorskiej nieinwazyjnej metody diagnozowania cukrzycy typu 1 opartej na pomiarze aktywności fizycznej, w której wykorzystano algorytmy inteligencji obliczeniowej, analizując główne parametry, takie jak tygodniowa liczba kroków i czas intensywnej aktywności. Badania wykazały, że cukrzycę typu 1 można skutecznie diagnozować na podstawie tych parametrów, co ma istotne znaczenie ze względu na nieinwazyjność i elastyczność metody. Autorzy zaproponowali implementację tej techniki na urządzeniu mobilnym. Warto podkreślić rosnącą popularność i akceptację rozwiązań mobilnych w dziedzinie zdrowia, szczególnie wśród młodego pokolenia.

- II. Drałus, G., Mazur, D., & **Czmił, A.** (2021). *Automatic Detection and Counting of Blood Cells in Smear Images Using RetinaNet*. *Entropy*, 23(11), 1522. doi:10.3390/e23111522. IF (2021): 2,738, liczba punktów: 100, wkład: 33,33%.

Tytuł zadania: Opracowanie metody pozwalającej na automatyczne, jednoczesne rozpoznawanie i zliczanie czerwonych i białych krwinek oraz płytek krwi na podstawie zdjęć mikroskopowych z wykorzystaniem głębokich sieci neuronowych.

Artykuł ten prezentuje automatyczną metodę liczenia komórek krwi przy użyciu uczenia maszynowego opartego na konwolucyjnych sieciach neuronowych RetinaNet, odpowiadającą na wyzwania laboratoriów medycznych. Ręczne liczenie komórek jest pracochłonne i podatne na błędy, co może mieć poważne konsekwencje dla całego procesu diagnostycznej. Autorzy opisują dwa główne podejścia: tradycyjne metody i głębokie sieci neuronowe, prezentując rozwój aplikacji wykorzystującej RetinaNet do jednoczesnego wykrywania i liczenia komórek krwi. Wyniki badań wskazują, że dokładność modelu zależy od ustawionego progu ufności, a opracowana aplikacja umożliwia łatwą weryfikację wyników. Artykuł podkreśla korzyści tej metody, takie jak zmniejszenie obciążenia pracy w laboratoriach, minimalizacja błędów ludzkich i zwiększenie dokładności liczenia komórek, sugerując potencjał zastąpienia ręcznych procesów w diagnostyce medycznej. Badanie demonstruje skuteczność zastosowania uczenia głębokiego w automatyzacji procesu liczenia komórek krwi, przynosząc istotne ulepszenia w efektywności i dokładności w porównaniu z tradycyjnymi metodami.

III. **Czmił, A.**, Wroński, M., Czmił, S., Sochacka-Piętal, M., Ćmił, M., Gawor, J., Wołkowicz, T., Plewczyński, D., Strzałka, D., & Piętal, M. (2022). NanoForms: an integrated server for processing, analysis and assembly of raw sequencing data of microbial genomes, from Oxford Nanopore technology. PeerJ, 10, e13056. doi:10.7717/peerj.13056. IF (2022): 3,061, liczba punktów: 100, wkład: 10%.

Tytuł zadania: Opracowanie aplikacji umożliwiającej automatyzację procesu oceny, składania i identyfikacji sekwencji genomowych uzyskanych za pomocą nowych metod sekwencjonowania przy użyciu narzędzi korzystających z metod uczenia maszynowego.

Praca ta przedstawia serwer NanoForms, nowoczesne narzędzie do analizy danych genetycznych z sekwencjonowania Oxford Nanopore. Serwer, oparty na języku Python i systemie Django, umożliwia przetwarzanie dużych ilości danych genetycznych, szczególnie przydatne po zakończeniu eksperymentu. NanoForms jest dostępny dla użytkowników akademickich bezpłatnie, a jego wirtualizacja pozwala na obsługę 5-10 równoległych zadań. Interaktywny interfejs i łatwość obsługi wyróżniają NanoForms w porównaniu do innych serwerów, co jest korzystne dla użytkowników bez doświadczenia bioinformatycznego. Badanie skupia się na sekwencjonowaniu genomu *Bacillus subtilis*, prezentując praktyczne zastosowanie technologii Oxford Nanopore. Autorzy przeprowadzili dokładną analizę i porównanie różnych usług, testując serwery takie jak CGE, Enterobase, Galaxy Tools, EPI2ME, NanoPipe, Patric i NanoGalaxy. Każdy z tych serwerów posiada unikalne cechy i funkcjonalności. W tym kontekście, NanoForms wyróżnia się łatwością obsługi i interaktywnością, co jest szczególnie korzystne dla użytkowników bez doświadczenia bioinformatycznego. Chociaż publikacja skupia się głównie na aspektach technicznych i informatycznych, to jednak wnosi znaczny wkład naukowy, prezentując sekwencjonowanie genów na przykładzie *Bacillus subtilis*.

- IV. **Czmił, A.**, Kluska, J., & Czmił, S. (2023). GPR: A Python implementation of an extremely simple classifier based on fuzzy logic and gene expression programming. *SoftwareX*, 22, 101362. doi:10.1016/j.softx.2023.101362. IF (2022): 2,868, liczba punktów: 200, wkład: 33,33%.

Tytuł zadania: Implementacja w języku Python klasyfikatora opartego na logice rozmytej i programowaniu ekspresji genów, służącego do generowania wysoce interpretowalnych reguł rozmytych.

Artykuł prezentuje algorytm klasyfikacji GPR, bazujący na prostych zasadach rozmytych "if-then", generowanych automatycznie za pomocą metod Gene Expression Programming (GEP). Ten algorytm adresuje trudność równoważenia interpretowalności reguł z dokładnością klasyfikacji w projektowaniu klasyfikatorów. Przetestowany na 16 zestawach danych, GPR okazał się jednym z najlepszych klasyfikatorów pod względem obszaru pod krzywą ROC i dokładności klasyfikacji. Zaimplementowany w języku Python przy użyciu Deap, Geppy i NumPy, algorytm jest dostępny na GitHubie na licencji MIT. Wspiera interfejs Scikit-learn, oferując metody **fit()** i **predict()**. Autorzy podkreślają, że GPR jest zgodny z koncepcją eksplikowalnej sztucznej inteligencji, co sprawia, że jest łatwy w użyciu bez dużego zaangażowania użytkownika. Algorytm może być stosowany w różnych dziedzinach, w tym w analizie danych medycznych.

- V. **Czmił, A.** (2023). Comparative Study of Fuzzy Rule-Based Classifiers for Medical Applications. *Sensors*, 23(2), 992. doi:10.3390/s23020992. IF (2022): 3,847, liczba punktów: 100, wkład: 100%.

Tytuł zadania: Opracowanie narzędzia pozwalającego na eksperymentalne porównanie wybranych rozmytych algorytmów opartych na regułach do klasyfikacji danych medycznych.

Publikacja ta skupia się na zastosowaniu algorytmów opartych na logice rozmytej do wspierania decyzji medycznych w diagnostyce i prognozowaniu chorób. Autorka omawia w niej wyzwania związane z niejednoznacznymi objawami chorób i potrzebę zarządzania danymi pacjentów. Systemy wspomaganie decyzji medycznych (MDSS), zwłaszcza oparte na regułach rozmytych (FRBS), są analizowane pod kątem interpretowalności dla ekspertów medycznych. Porównano różne algorytmy FRBS, zwracając uwagę na ich efektywność, długość i zrozumiałość generowanych reguł. GPR (klasyfikator oparty na logice rozmytej i programowaniu ekspresji genów) wyróżnia się krótkimi i zrozumiałymi regułami przy utrzymaniu dobrej wydajności klasyfikacji. Testy statystyczne potwierdzają porównywalność skuteczności GPR z innymi algorytmami FRBS, ale zauważają prostotę generowanych reguł. Ostatecznie, artykuł sugeruje możliwości dalszych badań, w tym ocenę wymagań pamięciowych i czasu działania algorytmów oraz rozwój interfejsu użytkownika dla specjalistów medycznych ułatwiającego generowanie reguł.

4. Uwagi redakcyjne, krytyczne oraz pytania do pracy.

Ad. 4.

- W treści publikacji będących częścią cyklu oraz na rysunkach można zauważyć drobne błędy interpunkcyjne, literówki, powtórzenia oraz ewidentne pomyłki edytorskie. Niemniej jednak, nie mają one wpływu na pozytywny odbiór całej pracy czy jej wartość merytoryczną.
- Czy zastosowanie głębokich sieci neuronowych zamiast prostych modeli, takich jak drzewa decyzyjne czy logika rozmyta, mogłoby być konkurencyjne w przedstawionych zadaniach badawczych? Doktorantka skorzystała z sieci głębokiej tylko w jednym zadaniu; jednak, czy można byłoby zastosować podobne rozwiązania w pozostałych zadaniach?
- Zadania omówione w pracy są dość różne, czy doktorantka zastanawiała się nad połączeniem swoich rozwiązań w jedną całość? Nad stworzeniem uniwersalnego systemu diagnostycznego?
- Czy planowana jest implementacja kliniczna któregoś z rozwiązań/zadań?
- Czy zostały wykonane jakieś działania próbne z udziałem personelu medycznego do oceny obsługi systemu z siecią RetinaNet? Czy planowane są odpowiednie szkolenia dla osób niezwiązanych bezpośrednio z informatyką?
- Praca [A-3] - co wpłynęło na wybór tylko 3 typów komórek krwi?
- Brakuje osobnego, szczegółowego rozdziału poświęconego dyskusji.
- Cytowanie Cat GPT [9] jest zbyt techniczne i na poziomie cytowania Wikipedii.
- Brakuje rozwiniętego wprowadzenia literaturowego w części pierwszej pracy.

5. Podsumowanie.

Ad. 5. Uważam, że Doktorantka z powodzeniem zrealizowała wszystkie wymienione cele, począwszy od pozyskania danych medycznych, ich przetworzenia i analizy, aż po sformułowanie konkretnych zadań badawczych, selekcję, projektowanie i tworzenie narzędzi informatycznych do ich rozwiązania. To obejmowało także opracowanie oprogramowania, przeprowadzenie badań na danych rzeczywistych, interpretację uzyskanych wyników i sformułowanie wniosków. Dodatkowo nakreśliła kierunki dalszych badań, co oznacza, że ma w planach dalszy rozwój naukowy.

Przedstawione do oceny prace naukowe zostały już wcześniej zrecenzowane w procesie publikacyjnym. Całość rozprawy, rozumianej jako zbiór artykułów naukowych, ma w tym przypadku charakter wybitnie interdyscyplinarny, łączący informatykę, medycynę oraz inżynierię biomedyczną, ale bez najmniejszych wątpliwości mieści się w dyscyplinie **Informatyka Techniczna i Telekomunikacja**.

Omawiane zagadnienia są prawidłowo zdefiniowane i przedstawione, co świadczy o dojrzałości doktorantki. A zawartość poszczególnych prac stanowi w mojej opinii bardzo dużą wartość badawczą, praktyczną i aplikacyjną i może w przyszłości stanowić nieocenioną pomoc przy usprawnieniu diagnostyki medycznej. Bardzo wysoko oceniam również wkład pani **Czmił** w poszczególne publikacje.

Na szczególną uwagę zasługują ogólna działalność naukowa doktorantki. Wskazane w cyklu publikacyjnym prace mają bardzo wysokie współczynniki cytowania (IF) oraz punktację. Doktorantka otrzymała uwzględniając tylko cykl publikacyjny **14,988** IF oraz **570** punktów; a w całym swym dorobku jej IF wynosi **18,78** a liczba punktów aż **910**, co jest bardzo wysokim wynikiem na tym etapie kariery naukowej. Pragnę również podkreślić, że w całym swym dorobku pani Anna Czmił posiada aż dwie publikacje za **200** punktów.

Jeśli chodzi o rozpoznawalność doktorantki, posiada ona indeks Hirscha: **4** (Google Scholar), **3** (Scopus), **3** (Web of Science). Sumaryczna liczba cytowań wszystkich prac wynosi: **49** (Google Scholar), **38** (Scopus); **31** (Web of Science), co stanowi dobre wartości na tym etapie rozwoju naukowego. Spotkałam się z osobami rozpoczynającymi procedurę habilitacyjną mających podobne wartości współczynnika Hirscha.

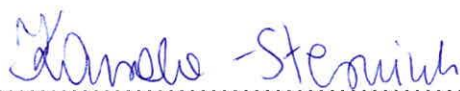
Cytowalność publikacji z cyklu jest następująca:

- [A-1]: **17** (Google Scholar), **14** (Scopus), **9** (Web of Science);
- [A-2]: **16** (Google Scholar), **11** (Scopus), **9** (Web of Science);
- [A-3]: **3** (Google Scholar), **3** (Scopus), **3** (Web of Science);
- [A-4]: **1** (Google Scholar), **1** (Scopus), **1** (Web of Science);
- [A-5]: **4** (Google Scholar), **2** (Scopus), **2** (Web of Science).

Ponadto doktorantka brała udział w **3** konferencjach naukowych, **2** projektach badawczych i jest laureatką **1** nagrody.

Moja ocena pracy **mgr inż. Anny Czmił** jest **pozytywna**. Moim zdaniem niniejsza praca prezentuje cenne wyniki badań i jest znaczącym osiągnięciem naukowym w dyscyplinie naukowej **Informatyka Techniczna i Telekomunikacja**. Spełnia ona również w mojej ocenie wszystkie wymogi zawarte w aktualnie obowiązującej Ustawie z dnia 20 lipca 2018 roku "Prawo o szkolnictwie wyższym i nauce" w sprawie warunków i trybu przeprowadzania przewodów doktorskich i może być przedmiotem publicznej obrony.

Wniosuję do Rady Naukowej Dyscypliny **Informatyka Techniczna i Telekomunikacja** o dopuszczenie pani **mgr inż. Anny Czmił** do dalszych etapów przewodu doktorskiego, a także o **wyróżnienie rozprawy**.



.....
Dr hab. Inż. Aleksandra Kawala-Sterniuk, prof. uczelni
Wydział Elektrotechniki, Automatyki i Informatyki
Politechnika Opolska
ul. Prószkowska 76
45-758 Opole
a.kawala-sterniuk@po.edu.pl