

000001353
Uniwersytet Marii Curie-Skłodowskiej
Wydział Matematyki, Fizyki i Informatyki
Katedra Neuroinformatyki i Inżynierii Biomedycznej
ul. Akademicka 9
20-033 Lublin, tel. 81 537-29-40

dr hab. Grzegorz Marcin Wójcik, prof. UMCS
Kierownik Katedry Neuroinformatyki i Inżynierii Biomedycznej
Uniwersytet Marii Curie-Skłodowskiej Lublinie
Instytut Informatyki
ul. Akademicka 9, 20-033 Lublin
gmwojcik@live.umcs.edu.pl

Recenzja rozprawy doktorskiej (sporządzonej w oparciu o cykl publikacji) mgr inż. Anny Czmił

Tytuł rozprawy: Usprawnienie procesu diagnostyki medycznej przy użyciu metod sztucznej inteligencji

Promotor w przewodzie: dr hab. inż. Damian Mazur, prof. PRz

Promotor pomocniczy: dr hab. n. med. Bogdan Obrzut, prof. UR

Rozwój sztucznej inteligencji (SI) stanowi dynamiczny obszar, który przekształca sposób, w jaki ludzie komunikują się, pracują i podejmują decyzje. W ciągu ostatnich kilku lat zaobserwowano niezwykle postęp w dziedzinie uczenia maszynowego, głębokiego uczenia i przetwarzania języka naturalnego. Algorytmy SI zdolne są teraz do skomplikowanej analizy dużych zbiorów danych, wykrywania wzorców i podejmowania decyzji na podstawie informacji, które dla ludzi mogą być trudne do zauważenia lub zrozumienia. Rozwój systemów SI obejmuje również szereg dziedzin, takich jak autonomiczne pojazdy, medycyna, finanse, czy robotyka. Technologie te znajdują zastosowanie w codziennym życiu, od asystentów w smartfonach po zaawansowane systemy diagnozowania chorób. Wraz z postępem w badaniach nad SI, naukowcy i inżynierowie zdobywają nowe umiejętności w tworzeniu inteligentnych systemów, co otwiera drzwi do nieograniczonych możliwości wykorzystania SI we wszystkich dziedzinach życia. Pomimo

WPŁYNEŁO

43 STY. 2024



korzyści, rozwój SI stawia również przed społeczeństwem wyzwania związane z etyką, prywatnością i bezpieczeństwem, wymagając równocześnie odpowiedzialnego podejścia do implementacji tych zaawansowanych technologii.

SI odgrywa coraz istotniejszą rolę w diagnostyce medycznej, przynosząc nowe, innowacyjne rozwiązania i poprawiając precyzję oraz efektywność procesów diagnostycznych. Algorytmy oparte na SI są zdolne do analizy ogromnych zbiorów danych medycznych w krótkim czasie, umożliwiając szybką identyfikację wzorców i wykrywanie subtelných anomalii, które mogą być trudne do zauważenia przez ludzki wzrok. W medycynie obrazowej, takiej jak tomografia komputerowa czy rezonans magnetyczny, SI wspomaga radiologów w analizie obrazów, pomagając w wykrywaniu i klasyfikowaniu zmian patologicznych. Ponadto, w dziedzinie analizy danych laboratoryjnych, SI może przewidywać ryzyko wystąpienia chorób na podstawie wyników badań krwi czy moczu. Wykorzystanie sztucznej inteligencji w diagnostyce medycznej nie tylko skraca czas oczekiwania na wyniki, ale także podnosi poziom dokładności diagnoz, co z kolei przyczynia się do szybszego i bardziej skutecznego leczenia pacjentów.

Niniejsza praca składa się z pięciu artykułów naukowych opublikowanych w latach 2019-2023 w języku angielskim, zgodnie z wymogami zawartymi w art. 187 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce. Publikacje skupiające się w ramach tego cyklu zostały wcześniej poddane recenzji w procesie publikacyjnym, prezentując innowacyjne rozwiązania związane z tematyką pracy doktorskiej. Celem pracy jest zbadanie i doskonalenie procesu diagnostyki medycznej, wykorzystując nowoczesne metody sztucznej inteligencji.

Tematyka rozprawy wypełnia założenia ustawowe wymagane do ubiegania się o stopień naukowy doktora w dziedzinie nauk technicznych, w dyscyplinie informatyka techniczna i telekomunikacja.

W ramach przedłożonej rozprawy, czyli de facto cyklu publikacji, autorka formułuje hipotezę badawczą stanowiącą, iż:

„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 weryfikacji hipotezy badawczej sformułowanych zostaje 5 celów badawczych:

1. Zastosowanie metod sztucznej inteligencji do klasyfikacji cukrzycy typu 1 na podstawie danych uzyskanych za pomocą nieinwazyjnych pomiarów aktywności fizycznej.
2. 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.
3. Opracowanie aplikacji umożliwiającej automatyzację procesu oceny, składowania i identyfikacji sekwencji genomowych uzyskanych za pomocą nowych metod sekwencjonowania przy użyciu narzędzi korzystających z metod uczenia maszynowego.
4. Implementacja w języku Python klasyfikatora opartego na logice rozmytej i programowaniu ekspresji genów, służącego do generowania wysoce interpretowalnych reguł rozmytych.
5. Opracowanie narzędzia pozwalającego na eksperymentalne porównane wybranych rozmytych algorytmów opartych na regułach do klasyfikacji danych medycznych.

Przedłożony do recenzji cykl artykułów stanowiący rozprawę obejmuje 146 stron, na których znajdujemy:

1. Część wprowadzającą do dorobku naukowego autorki z opisem wybranych metod SI wykorzystywanych dalej, podsumowaniem oraz spisem literatury, także z wykazem skrótów i oznaczeń w mojej ocenie bardzo potrzebnych podczas lektury tekstu jako całości (38 stron).
2. Artykuły naukowe opublikowane w latach 2019-2023 – stanowiące główną część rozprawy (82 strony).
3. Streszczenia w językach polskim oraz angielskim.

Dodatkowo zamieszczono oświadczenia współautorów prac doktorantki, które potwierdzają jej istotny udział w projekcie.

Struktura pracy jak i rozkład treści są w mojej ocenie poprawne.

Omawiane pięć artykułów stanowią prace realizujące kolejno każdy z pięciu wyznaczonych przez autorkę i przytoczonych w niniejszej recenzji celów:

1. **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)**

Praca skupia się na innowacyjnej metodyce nieinwazyjnej diagnozy cukrzycy typu 1, opartej na monitorowaniu aktywności fizycznej i wykorzystującej algorytmy inteligencji obliczeniowej do analizy kluczowych parametrów, takich jak tygodniowa liczba kroków czy czas intensywnej aktywności. Zbiór danych obejmował grupę 230 dzieci w wieku od 6 do 18 lat, w tym 115 dzieci zdrowych oraz 115 dzieci z cukrzycą typu 1, będących pod opieką Poradni Cukrzycowej dla dzieci w Klinicznym Szpitalu Wojewódzkim nr 2 w Rzeszowie. Zastosowano metody klasyfikacji binarnej,

określenia współczynnika korelacji oraz klastrowania. Oceny jakości klasyfikacji binarnej dokonano za pomocą metryk wydajnościowych, takich jak dokładność (accuracy, ACC), czułość (sensitivity, Sen), specyficzność (specificity, Spe), precyzja (precision, Pre), wskaźnik dobroci (goodness index, G-index) i pole pod krzywą charakterystyki działania odbiornika (area under ROC curve, AUC). Najlepsze wyniki dokładności (86,09%), specyficzności (84,35%), precyzji (84,87%) oraz wskaźnika zgodności (0,1983) uzyskano, stosując metodę lasu drzew decyzyjnych. Największą czułość osiągnęła probabilistyczna sieć neuronowa – 89,57%.

2. **Drałus, G., Mazur, D., & Czmił, A. (2021). Automatic Detection and Counting of Blood Cells in Smear Images Using RetinaNet. Entropy, 23(11), 1522.**

W pracy autorzy koncentrują się na automatycznym zliczaniu komórek krwi z wykorzystaniem konwolucyjnych sieci neuronowych. Wykorzystywano model Retina-Net, który odpowiednio wytrenowany osiągał skuteczność powyżej 90%. Warto tu wskazać, że autorzy znaleźli optymalną liczbę epok dla modelu, w taki sposób, by osiągać najwyższą skuteczność i jednocześnie unikać przetrenowania.

3. **Czmił, A., Wroński, M., Czmił, S., Sochacka-Piętal, M., Ćmił, M., Gawor, J., Wołkiewicz, 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.**

Praca zawiera opis aplikacji serwerowej napisanej w języku Python/Django pozwalającej przeprowadzać analizę danych genetycznych pochodzących z Oxford Nanopore. Aplikacja powstała zgodnie ze najwyższymi standardami inżynierii oprogramowania, pozwalała na stabilną pracę do 10 użyt-

kowników. Dla użytkowników akademickich aplikacja udostępniana jest bezpłatnie. Wybór technologii oraz metodologii projektowania aplikacji poprzedzały dokładne analizy wydajnościowe wielu konkurencyjnych platform. Aplikacja, którą współtworzyła doktorantka cechuje się wysokim stopniem intuicyjności obsługi i interakcji.

4. **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.**

Praca zawiera implementację klasyfikatora GPR. W skrócie, służy on do generowania interpretowalnych reguł rozmytych w logice rozmytej z zastosowaniami w bioinformatyce. Zważywszy na dyscyplinę, w jakiej doktorantka zamierza ubiegać się o stopień doktora, należy podkreślić, że implementacja algorytmu opisanego w tym artykule integruje kilka ważnych bibliotek i narzędzi współczesnej inżynierii dużych zbiorów danych, w szczególności: Deap, Geppy, Scikit-learn, NumPy. Powyższe świadczą o dojrzałości autorki w zakresie zastosowania metod programowania, głównie zresztą w języku Python, który staje się wiodącym językiem współczesnej Data Science.

5. **Czmił, A. (2023). Comparative Study of Fuzzy Rule-Based Classifiers for Medical Applications. *Sensors*, 23(2), 992.**

O ile w wcześniejszych pracach udział autorki stanowił ok. 30%, o tyle w tym artykule udziały te wynoszą 100% i należy uznać go za najważniejszy z całego cyklu. Praca koncentruje się na wykorzystaniu algorytmów opartych na logice rozmytej w celu wspierania decyzji medycznych w diagnostyce i prognozowaniu chorób, co właściwie zawiera się w tytule rozprawy stanowiącej cykl publikacji. Autorka analizuje wyzwania związane z niejednoznacznymi objawami chorób oraz konieczność skutecznego

zarządzania danymi pacjentów. Przeanalizowano różne systemy wspomaganie decyzji medycznych (MDSS), ze szczególnym uwzględnieniem tych opartych na regułach rozmytych (FRBS), z myślą o interpretowalności dla ekspertów medycznych. Oceniono różne algorytmy FRBS (16 na 12 zbiorach danych), zwracając uwagę na ich skuteczność, długość oraz zrozumiałość generowanych reguł. Klasyfikator GPR (oparty na logice rozmytej i programowaniu ekspresji genów) wyróżnia się krótkimi i zrozumiałymi regułami, przy jednoczesnym utrzymaniu wysokiej wydajności klasyfikacji. Testy statystyczne potwierdzają porównywalność skuteczności GPR z innymi algorytmami FRBS, przy jednoczesnym zauważeniu prostoty generowanych reguł. Przeprowadzono poprawną analizę statystyczną. W podsumowaniu, artykuł wskazuje na możliwości dalszych badań, w tym ocenę wymagań pamięciowych i czasu działania algorytmów, a także rozwój interfejsu użytkownika dla specjalistów medycznych, co ułatwi dalsze generowanie reguł.

Pytania i uwagi krytyczne

Niewdzięcznym obowiązkiem recenzenta jest za to wytknięcie spostrzeżonych uchybień i uwag technicznych.

Podczas lektury rozprawy nie zauważyłem uchybień, które miałyby jakiś szczególnie istotny wpływ na jej jednoznacznie pozytywną ocenę.

Natomiast pojawiło się u mnie kilka pytań, do których chciałbym by doktorantka ustosunkowała się w przypadku dopuszczenia do obrony. Pytania te wynikają z ciekawości, nie zaś z zastrzeżeń do treści:

1. Czy można byłoby rozszerzyć któreś z zadań/rozwiązań o użycie dużych modeli językowych (LLM)?
2. Jak wygląda proces pozyskiwania informacji o pacjentach do systemów

mających na celu wspomaganie diagnozy? Jak wyglądają aspekty prawne?

3. Czy planowane jest rozwinięcie algorytmu GPR generującego reguły rozmyte?
4. Czy rozważane są wdrożenia chmurowe, jakie?

Wszystkie powyższe pytania i uwagi mają charakter pytań z natury dociekliwych i nie mają wpływu na moje jednoznacznie pozytywne wrażenie z lektury pracy doktorskiej stanowiącej cykl artykułów pani mgr inż. Anny Czmił.

Rekomendacja

Pani mgr inż. Anna Czmił zaczyna być zauważalną, młodą badaczką. Według Google Scholar jej indeks Hirscha wynosi 4, przy liczbie cytowań 49. W innych bazach, jak pokazuje doświadczenie, wskaźniki będą również wzrastać, co wymaga jednak czasu, a badania i prace, w których zostały opisane są przecież relatywnie nowe.

Moja ocena rozprawy doktorskiej p. mgr. inż. Anny Czmił jest zdecydowanie pozytywna.

Uważam, że rozprawa doktorska mgr. inż. Anny Czmił spełnia warunki określone w art. 187 Ustawy z dnia 20 lipca 2018 r. o szkolnictwie wyższym i nauce (Dz. U. 2018 poz. 1668), dlatego zwracam się do Wysokiej Rady Dyscypliny Informatyki Technicznej i Telekomunikacji Wydziału Elektrotechniki i Informatyki o dopuszczenie mgr. inż. Anny Czmił do dalszych etapów przewodu doktorskiego. Mając na uwadze wysoce interdyscyplinarny charakter rozprawy oraz liczbę opublikowanych przez doktorantkę prac wnoszę o jej wyróżnienie.

Lublin, 2024-01-07


Kierownik Katedry
dr hab. Grzegorz M. Wójcik
prof. UMCS