

Prof. Marek Wesółowski
Katedra i Zakład Chemii Analitycznej
Gdański Uniwersytet Medyczny

Recenzja

pracy doktorskiej mgr Andżeliki Lorenc zatytułowanej *Zastosowanie algorytmów uczenia maszynowego w analizie predykcyjnej aktywności przeciwdrobnoustrojowej pochodnych imidazoliowych i wznowy raka płuca,*

wykonanej w Katedrze Biofarmacji Wydziału Farmaceutycznego Collegium Medicum w Bydgoszczy, Uniwersytet Mikołaja Kopernika w Toruniu

**Promotorzy: prof. dr hab. n. farm. Adam Buciński
dr hab. n. med. Anna Romaszko-Wojtowicz, prof. UWM**

Przedstawiona do oceny rozprawa doktorska mgr Andżeliki Lorenc wychodzi naprzeciw aktualnym trendom związanym z zastosowania sztucznej inteligencji, w tym uczenia maszynowego, jako narzędzi nadzwyczaj użytecznych w rozwiązywaniu różnych problemów badawczych, odznaczających się szczególnie dużym stopniem złożoności. Z tego względu, dogłębna wiedza i szczególne umiejętności w zakresie budowania i wdrażania modeli uczenia maszynowego są bardzo pożądanymi kompetencjami, zarówno w badaniach naukowych, jak i w dynamicznie rozwijającej się technologii informacyjnej. O dużym stopniu złożoności tego zagadnienia świadczy m.in. fakt, że problematyka uczenia maszynowego obejmuje nie tylko podstawowe uczenie maszynowe, ale też zaawansowane techniki uczenia głębokiego. Dlatego zajęcie się problematyką uczenia maszynowego wymaga odpowiednich kompetencji nie tylko w zakresie wiedzy o przygotowaniu wielkich zbiorów danych, ale także w kwestii wyboru i trenowania modeli sieci neuronowych, algorytmów klasyfikacji i regresji, drzew decyzyjnych oraz technik redukcji wymiarowości.

Biorąc powyższe pod uwagę, realizowane przez mgr Andżelikę Lorenc szeroko zakrojone badania nad zastosowaniem zaawansowanych algorytmów uczenia maszynowego w analizie predykcyjnej aktywności przeciwdrobnoustrojowej pochodnych imidazoliowych i wznowy raka płuca, są nadzwyczaj aktualne i wielce pożądane oraz szczególnie użyteczne z praktycznego punktu widzenia. Umiejętność modelowania złożonych zjawisk biologicznych i klinicznych umożliwiła bowiem Doktorantce skuteczne odróżnienie związków aktywnych przeciw drobnoustrojom od związków nieaktywnych, a ponadto identyfikację najważniejszych czynników predykcyjnych wznowy raka płuca, co może mieć bardzo duże znaczenia przy

podejmowaniu decyzji klinicznych.

Dysertacja doktorska mgr Andżeliki Lorenc ma układ typowy. Składa się z pięciu ponumerowanych rozdziałów, w tym *Wstępu*, *Celu pracy*, *Materiałów i metod*, *Wyników* oraz *Podsumowania i wniosków*. W jej skład wchodzi również *Literatura* i *Spis rycin*. Właściwą treść rozprawy poprzedza strona tytułowa, podziękowania, spis treści oraz wykaz skrótów i akronimów. Istotnym elementem są także zwięzłe, jednostronicowe streszczenia, w języku polskim i angielskim. Pisząc ten bardzo ważny fragment każdej pracy naukowej, Doktorantka wykazała się umiejętnością zawarcia w krótkim tekście najważniejszych osiągnięć własnych, kilkuletnich badań, co wcale nie jest łatwe. Ponadto, do treści rozprawy włączono oświadczenie Komisji Bioetycznej, opis merytorycznego udziału Doktorantki w publikacjach, oświadczenia współautorów oraz życiorys naukowy i wykaz publikacji Doktorantki. Całości dopełnia zbiór trzech powiązanych tematycznie artykułów przedstawiających wyniki badań wykonanych w ramach pracy doktorskiej.

Realizację zasadniczego celu pracy mgr Andżelika Lorenc poprzedziła liczącym 17 stron *Wstępem*, w którym udowodniła, że bardzo dobrze zrozumiała poszczególne kwestie związane z uczeniem maszynowym i sztucznymi sieciami neuronowymi oraz z uzyskaniem satysfakcjonujących wyników. Jest więc bardzo dobrze przygotowana do rozwiązywania problemów, z jakimi spotka się podczas realizacji celu badań, a następnie interpretacji i publikacji osiągniętych wyników. We *Wstępie* przygotowanym na podstawie 57 pozycji literatury, Doktorantka w zwięzły sposób wyjaśniła pojęcia sztucznej inteligencji i uczenia maszynowego, a ponadto zilustrowała relacje hierarchiczne pomiędzy tymi pojęciami oraz sztucznymi sieciami neuronowymi i uczeniem głębokim. Scharakteryzowała również krótko typy uczenia maszynowego (nadzorowane i nienadzorowane) oraz wybrane algorytmy uczenia maszynowego, tj. drzewa decyzyjne (drzewa klasyfikacyjne i regresyjne, CART), analizę składowych głównych (PCA) i sztuczne sieci neuronowe (ANN).

Szczególne znaczenie ANN jako narzędzi do realizacji celu badań też znalazło swoje odzwierciedlenie we *Wstępie*. Omawiając parametry architektury i uczenia sieci neuronowej, Doktorantka w wyczerpującym stopniu scharakteryzowała algorytm wstecznej propagacji błędów, najlepiej poznany i najczęściej stosowany do uczenia sieci oraz metody gradientów sprzężonych i quasi-Newtonowskie. Omówiła również znaczenie podziału zbioru danych na zbiory uczący, walidacyjny i testowy oraz wyjaśniła pojęcia generalizacji i przeuczenia wraz z charakterystyką strategii ograniczających przeuczenie i prowadzących do poprawy zdolności generalizacji. Zagadnienia omówione we *Wstępie* wskazują, że odpowiedni dobór architektury

sieci, funkcji aktywacji i procedury optymalizacji parametrów zapewnia minimalizację funkcji straty oraz dopasowanie modelu do obserwowanych danych na poziomie umożliwiającym implementowanie go do nowych danych.

Cel pracy doktorskiej został sformułowany jasno i logicznie, z wyszczególnieniem czterech celów szczegółowych. Nadrzędnym zadaniem mgr Andżeliki Lorenc była ocena możliwości wykorzystania wybranych metod uczenia maszynowego, tj. CART, PCA i ANN typu perceptron wielowarstwowy (MLP) do modelowania i predykcji złożonych problemów medycznych. Cele szczegółowe 1 i 2 polegały na opracowaniu i walidacji modeli ANN typu MLP do klasyfikacji aktywności przeciwdrobnoustrojowej wobec *Klebsiella pneumoniae* i ilościowej predykcji wartości logarytmu naturalnego z minimalnego stężenia hamującego wobec *Staphylococcus aureus* oraz *Candida albicans* dla odpowiednio, 140 i 126 nowo syntetyzowanych pochodnych imidazoliowych. Istota badań polegała na selekcji najlepszych deskryptorów molekularnych opisujących zoptymalizowane struktury badanych związków za pomocą metody CART, redukcja wielowymiarowości metodą PCA oraz ocena zdolności do generalizacji opracowanych modeli.

Cel szczegółowy 3 polegał z kolei na opracowaniu i ocenie modelu ANN typu MLP do predykcji wznowy raka płuca w oparciu o retrospektywne dane kliniczne obejmujące 2296 pacjentów hospitalizowanych w Warmińsko-Mazurskim Centrum Chorób Płuc w Olsztynie (lata 2012-2017). Istota badania sprowadzała się do doboru zmiennych predykcyjnych o znaczeniu klinicznym, optymalizacji architektury sieci neuronowej i algorytmów uczenia oraz do analizy jakości klasyfikacyjnej. Cel szczegółowy 4 zmierzał natomiast w kierunku analizy porównawczej skuteczności użytych metod uczenia maszynowego w modelowaniu wielowymiarowych relacji pomiędzy danymi wejściowymi a zmienną wyjściową, na podstawie analizy danych o różnym charakterze, tj. molekularnym i klinicznym, biorąc pod uwagę wpływ wysokiej wymiarowości danych na proces uczenia, rolę selekcji zmiennych i redukcji wielowymiarowości w budowie stabilnych modeli oraz znaczenie procedur walidacyjnych w ocenie zdolności do generalizacji.

Wykonany przez Doktorantkę zakres prac obliczeniowych związanych z realizacją celu badań, dokonanych przy użyciu specjalistycznych programów komputerowych i korzystając z różnych baz danych, wraz z uzyskanymi wynikami, ich interpretacją, dyskusją i wnioskami, został opublikowany w formie trzech artykułów naukowych:

1. Lorenc A., Badura A., Karolak M., Pałkowski Ł., Kubik Ł., Buciński A.: Antimicrobial activity classification of imidazolium derivatives predicted by artificial

- neural networks, *Pharmaceutical Research*, 2024, 41, 891-898, <https://doi.org/10.1007/s11095-024-03699-x> (IF = 4,3; punkty ministerstwa = 70),
2. Lorenc A., Badura A., Karolak M., Pałkowski Ł., Kubik Ł., Skrzypczak A., Buciński A.: Hybrid machine learning model for dual prediction of antimicrobial activity against *Staphylococcus aureus* and *Candida albicans*. *Journal of Pharmaceutical Innovation*, 2026, 21, 174, <https://doi.org/10.1007/s12247-026-10412-5> (IF = 2,7; punkty ministerstwa = 70),
 3. Lorenc A., Romaszko-Wojtowicz A., Jaśkiewicz Ł., Doboszyńska A., Buciński A.: Exploring the efficacy of artificial neural networks in predicting lung cancer recurrence: a retrospective study based on patient records. *Translational Lung Cancer Research*, 2023, 12, 2083-2097, <https://dx.doi.org/10.21037/ticr-23-350> (IF = 4,0; punkty ministerstwa = 140).

Publikacja wyników wieloletnich badań wchodzących w zakres pracy doktorskiej w recenzowanych czasopismach o obiegu międzynarodowym, z których dwa zaliczane są do kwartyła Q1, świadczy o wysokim poziomie naukowym badań, poprawności merytorycznej i zainteresowaniu tym tematem. Łączna wartość współczynnika IF dla tych prac wynosi 11,0, natomiast punktacja ministerstwa = 280. Wszystkie artykuły opublikowano w czasopismach przypisanych do dyscypliny nauki farmaceutyczne. We wszystkich Doktorantka jest zarówno pierwszym, jak i korespondencyjnym autorem. Zgodnie z oświadczeniami, Doktorantka jest również twórcą koncepcji dwóch prac i współtwórcą koncepcji trzeciej, a ponadto projektowała, wykonywała i interpretowała wyniki obliczeń z użyciem algorytmów uczenia maszynowego, tj. CART, PCA i MLP oraz przygotowała manuskrypty do druku. Potwierdza to wiodący udział mgr Andżeliki Lorenc w uzyskaniu wartościowych wyników badań.

W efekcie badań zrealizowanych w ramach pracy doktorskiej, Doktorantka opracowała i dokonała walidacji modeli ANN typu MLP, zdolnych do klasyfikacji aktywności przeciwdrobnoustrojowej wobec *Klebsiella pneumoniae* (publikacja nr 1) oraz ilościowej predykcji wartości logarytmu naturalnego z minimalnego stężenia hamującego wobec *Staphylococcus aureus* i *Candida albicans* (publikacja nr 2) dla nowo syntetyzowanych pochodnych imidazoliowych. Do szczególnych osiągnięć należy zaliczyć opracowanie wieloetapowego procesu przetwarzania zbioru danych (selekcji deskryptorów molekularnych), wiążącego się z redukcją wielowymiarowości i eliminacją współliniowości deskryptorów oraz opracowanie formy uczenia wielozadaniowego, która polega na użyciu pojedynczego modelu sieci do wyznaczenia dwóch biologicznie odrębnych problemów predykcyjnych. Osiągnięcia te należy rozpatrywać w kategorii nowości metodologicznej, podobnie jak opracowaną w publikacji nr 3 procedurę doboru predyktorów w przypadku retrospektywnych danych klinicznych użytych

do predykcji wznowy raka płuca.

Najważniejsze osiągnięcia badań mgr Andżelika Lorenc przedstawiła w rozdziale *Podsumowanie i wnioski*. Wskazała przede wszystkim na fakt, że skuteczność ANN typu MLP zależy przede wszystkim od charakteru danych i sposobu ich przygotowania. Udowodniła to stosując drzewa decyzyjne CART do selekcji najistotniejszych zmiennych, a następnie PCA do redukcji wielowymiarowości, redukując tym sposobem liczbę zmiennych wejściowych do kilku, wyjaśniających ok. 99% całkowitej wariancji. Efektem tych działań była poprawa stabilności modeli i ich zdolności do generalizacji. Odmienna procedura optymalizacji poprzez użycie algorytmu wstecznej propagacji błędu oraz metody gradientów sprzężonych pozwoliła również na poprawę stabilności uczenia modelu i skuteczną minimalizację funkcji błędu w przypadku danych klinicznych, o mniejszej liczbie zmiennych, większej heterogeniczności i niezrównoważeniu klas. Końcowym efektem badań Doktorantki było kluczowe stwierdzenie, że ANN zachowują wysoką skuteczność predykcyjną pod warunkiem dostosowania procedur przygotowania danych i walidacji modeli do specyfiki analizowanego problemu.

Podsumowując stwierdzam, że cel pracy doktorskiej został zrealizowany. Badania są nowatorskie, stanowią oryginalne rozwiązania problemów naukowych i mogą być użyteczne w praktyce. Potwierdzają posiadanie przez Doktorantkę wymaganej wiedzy teoretycznej i możliwości jej praktycznego wykorzystania. Badania wymagające systematyczności i bardzo dużego nakładu pracy zostały precyzyjnie zaplanowane, a sposób prezentacji wyników, ich interpretacji i dyskusji świadczy z kolei o dojrzałości naukowej Doktorantki i umiejętności samodzielnego prowadzenia prac naukowych.

Piśmiennictwo ujęte w rozprawie doktorskiej obejmuje 59 profesjonalnie dobranych, angielskojęzycznych publikacji i monografii. Należy podkreślić aktualność zebranej literatury, więcej niż dwie trzecie prac zostało opublikowane po roku 2000. Literatura jest kolejnym elementem, który potwierdza bardzo dobre przygotowanie teoretyczne Doktorantki do prowadzenia badań w obszarze uczenia maszynowego, poprawnej interpretacji uzyskanych danych i ich krytycznej dyskusji. Drobne usterki w piśmiennictwie to pełne imiona autorów zamiast inicjałów (poz. 10 i 19), brak konsekwencji w zapisie tytułów czasopism – podawano zamiennie pełne nazwy albo skróty czasopism oraz niepełne dane bibliograficzne, szczególnie w przypadku monografii.

Pod względem redakcyjnym rozprawa doktorska została opracowana bardzo starannie, a układ pracy jest przejrzysty i logiczny. Wydaje się jednak, że Doktorantka zamieniła abstrakty graficzne na str. 35 i 37. Podpis pod Ryciną 5 odpowiada treści publikacji nr 1, a graficzny

abstrakt na tej samej stronie ilustruje badania zrealizowane w publikacji nr 2. To samo dotyczy podpisu pod Ryciną 6, odpowiada treści publikacji nr 2, natomiast abstrakt graficzny ilustruje badania z publikacji nr 1. Moją wątpliwość wzbudza także fakt umieszczenia podrozdziałów 3.2. Ocena jakości modeli klasyfikacyjnych oraz 3.3. Ocena jakości modeli regresyjnych w rozdziale *Materiały i metody*. Uważam, że powinny znajdować się we *Wstępie*.

Należy też wspomnieć, że mgr Andżelika Lorenc jest współautorką łącznie 11 artykułów opublikowanych w czasopismach zagranicznych i polskich (2 prace), o łącznym IF = 83,503 i punktacji ministerstwa = 1190. Uczestniczyła w czterech projektach badawczych, licznych kursach i szkoleniach, a za działalność naukową i dydaktyczną została wyróżniona nagrodami JM Rektora Uniwersytetu Mikołaja Kopernika w Toruniu.

W świetle wyżej przedstawionej, wysokiej oceny dysertacji doktorskiej autorstwa mgr A. Lorenc nt. *Zastosowanie algorytmów uczenia maszynowego w analizie predykcyjnej aktywności przeciwdrobnoustrojowej pochodnych imidazoliowych i wznowy raka płuca*, stwierdzam, że praca doktorska spełnia warunki formalne i merytoryczne stawiane rozprawom doktorskim w art. 187 Ustawy z dnia 20 lipca 2018 r., Prawo o Szkolnictwie Wyższym i Nauce (Dz.U. z 2024 r. poz. 1571). W związku z powyższym wnoszę do Wysokiej Rady Dyscypliny Nauki Farmaceutyczne w Collegium Medicum im. Ludwika Rydygiera w Bydgoszczy, Uniwersytetu Mikołaja Kopernika w Toruniu, o dopuszczenie mgr Andżeliki Lorenc do dalszych etapów postępowania w sprawie nadania stopnia doktora w dziedzinie nauk medycznych i nauk o zdrowiu, dyscyplinie nauki farmaceutyczne. Składam również wniosek o wyróżnienie rozprawy doktorskiej.

Gdańsk, dnia 4 września 2026 r.

Gdański Uniwersytet Medyczny
Katedra i Zakład Chemii Analitycznej

Prof. dr hab. Marek Wesółowski