

OCENA

rozprawy doktorskiej mgr farm. Łukasza Ledzińskiego pt. "Implementacja metod sztucznej inteligencji dla ułatwienia podejmowania decyzji klinicznych w kardiologii" wykonanej pod kierunkiem prof. dr hab. n. med. Grzegorza Grzeška z Katedry Kardiologii i Farmakologii Klinicznej, Wydziału Farmaceutycznego Collegium Medicum Uniwersytetu Mikołaja Kopernika w Toruniu.

Metody sztucznej inteligencji coraz częściej znajdują zastosowanie w medycynie, stanowiąc wsparcie lekarzy w analizie złożonych danych i podejmowaniu decyzji klinicznych. Dzięki algorytmom uczenia maszynowego możliwe jest szybkie przetwarzanie dużych zbiorów informacji, np. wyników badań laboratoryjnych, obrazów diagnostycznych oraz danych klinicznych pacjenta. Sztuczna inteligencja może pomagać w identyfikacji wzorców, które są trudne do zdefiniowania w tradycyjnej analizie prowadzonej wyłącznie przez człowieka. Szczególne znaczenie mogą odgrywać systemy wspomagania decyzji klinicznych, które stanowią narzędzie ułatwiające podejmowanie bardziej trafnych i szybszych decyzji przez lekarza: mogą sugerować możliwe rozpoznania, oceniać ryzyko powikłań lub wskazywać optymalne metody leczenia, w tym dostosowanie leczenia do indywidualnych cech pacjenta. Ważnym obszarem zastosowań sztucznej inteligencji jest także analiza obrazów medycznych, na przykład zdjęć rentgenowskich, tomografii komputerowej i rezonansu magnetycznego. Rozwój tych technologii wiąże się jednak z koniecznością zachowania przejrzystości działania algorytmów oraz odpowiedzialności za podejmowane decyzje. Sztuczna inteligencja w medycynie powinna być traktowana jako instrument wsparcia dla specjalistów, a nie jako samodzielny podmiot decyzyjny. Dlatego też próba implementacji metod sztucznej inteligencji dla ułatwienia podejmowania decyzji klinicznych w kardiologii jest jak najbardziej uzasadniona.

Na rozprawę dokorską mgr farm. Łukasza Ledzińskiego składają się trzy opublikowane prace, tj. 2 prace oryginalne: Ledziński Ł, Grzešek E, Ledzińska M, Grzešek G. Predicting Mortality in Atrial Fibrillation Patients Treated with Direct Oral Anticoagulants: A Machine Learning Study Based on the MIMIC-IV Database. *Journal of Clinical Medicine* 2025; 14(11):3697 (IF=2,9) i Ledziński Ł, Grzešek G, Ziołkowski M, Waligóra M, Kurzyna M, Mularek-Kubzdela T, Smukowska-Gorynia A, Skoczylas I, Chrzanowski Ł, Błaszczak P, et al. Personalized Medicine in Pulmonary Arterial Hypertension: Utilizing Artificial Intelligence for Death Prevention. *Journal of Clinical Medicine* 2025; 14(23):8325 (IF=2,9) oraz praca przeglądowa – Ledziński Ł, Grzešek G. Artificial Intelligence Technologies in Cardiology. *Journal of Cardiovascular Development and Disease* 2023; 10(5):202 (IF=2,3). We wszystkich powyższych publikacjach Doktorant jest pierwszym autorem, jego udział w powstaniu wymienionych prac wynosi 20-60%, co zostało potwierdzone przez

pozostałych współautorów poszczególnych prac. Ponadto przedłożona do oceny praca zawiera wstęp, omówienie prac składających się na pracę doktorską obejmujące przedstawienie celu pracy, materiałów i metody, wyników, dyskusji oraz wniosków, jak również streszczeń w języku polskim i angielskim, piśmiennictwo, kopie publikacji, oświadczenia współautorów oraz wykazy rycin, równań i tabel. Piśmiennictwo jest adekwatnie dobrane i cytowane.

We wstępie Doktorant wskazał na dynamiczny rozwój danych medycznych oraz ograniczenia wynikające z zastosowania klasycznych metod ich analizy wobec skali, różnorodności i tempa napływu informacji, stąd też niezbędne jest wprowadzenie nowych bardziej wydajnych narzędzi do analizy złożonych danych oraz wspierania decyzji klinicznych, takich jak sztuczna inteligencja (AI), uczenie maszynowe (ML) oraz duże modele językowe (LLM). Szczególną dziedziną potencjalnej implementacji ww. narzędzi jest kardiologia, w tym migotanie przedsionków (AF) i tętnicze nadciśnienie płucne (PAH), ze względu na dużą częstość chorób sercowo-naczyniowych, złożoność diagnostyki oraz wielość źródeł danych klinicznych. Doktorant pokazuje, że obecne skale stratyfikacji ryzyka dotyczące AF i PAH mają ograniczenia, co dobrze wskazuje na potrzebę zastosowania metod uczenia maszynowego. Całościowo wstęp jest merytorycznie spójny, aktualny i dobrze uzasadnia cel rozprawy którym jest wykorzystanie uczenia maszynowego do predykcji śmiertelności u pacjentów z PAH i AF. W treści wstępu być może należałoby wprowadzić pewne uszczegółowienia dotyczące skali CHA₂DS-VASc, która została zmodyfikowana przez ESC (2024) do CHA₂DS₂-VA, bez kategorii płci (Sc), jednakże w USA wytyczne ACC/AHA/HRS 2023 nadal szeroko posługują się CHA₂DS₂-VASc.

W oparciu o przesłanki zawarte we wstępie oraz pracy poglądowej (*Journal of Cardiovascular Development and Disease* 2023; 10(5):202) Doktorant sformułował następujące cele badawcze obejmujące: 1. cel główny: opracowanie modelu predykcyjnego z wykorzystaniem algorytmów sztucznej inteligencji, wspierającego codzienną praktykę kliniczną w opiece nad pacjentami z chorobami sercowo-naczyniowymi oraz 2. cele szczegółowe: zbadanie możliwości zastosowania algorytmów uczenia maszynowego do przewidywania ryzyka zgonu pacjentów oraz identyfikacji kluczowych czynników wpływających na predykcje modeli w grupie pacjentów chorujących na migotanie przedsionków leczonych doustnymi lekami przeciwkrzepliwymi oraz w grupie pacjentów chorujących na płucne nadciśnienie tętnicze w populacji krajowej.

Dla zrealizowania postawionych celów Doktorant zastosował algorytmy uczenia maszynowego oparte na drzewach decyzyjnych, co jest uzasadnione w analizach klasyfikacyjnych i regresyjnych danych klinicznych. Przedstawiona metodyka jest dobrze dobrane do celu rozprawy, ponieważ łączy klasyczną analizę statystyczną z nowoczesnymi metodami uczenia maszynowego stosowanymi w predykcji klinicznej. Doktorant zastosował algorytmy oparte na drzewach decyzyjnych, co jest uzasadnione w analizach

klasyfikacyjnych i regresyjnych danych klinicznych. Autor trafnie zauważył, że pojedyncze drzewa decyzyjne, mimo dużej przejrzystości i łatwości interpretacji, są podatne na nadmierne dopasowanie do danych uczących. Z tego względu uzasadnione było zastosowanie metod zespołowych, takich jak las losowy (ang. *Random Forest*, RF) oraz algorytmy wzmacniania, w tym XGBoost, LightGBM i AdaBoost. Trafne jest także uwzględnienie konieczności dostrajania hiperparametrów, zwłaszcza w przypadku modeli opartych na wzmacnianiu, które mogą osiągać wysoką skuteczność, ale wymagają starannej optymalizacji. Publikacja - *Journal of Clinical Medicine* 2025; 14(11):3697 opierała się na retrospektywnych danych z bazy MIMIC-IV w wersji 3.1, obejmującej zanonimizowane dane 364 627 pacjentów z lat 2008–2022. Do badania wybrano pacjentów z migotaniem przedsionków, którzy rozpoczęli leczenie doustnymi antykoagulantami podczas hospitalizacji; wykluczono chorych zmarłych w trakcie pobytu w szpitalu. Analiza obejmowała 600 zmiennych klinicznych, demograficznych, laboratoryjnych i terapeutycznych, a rozkład klas wynosił 6086:345. Zbiór podzielono na część uczącą i testową w proporcji 80:20, natomiast selekcja cech metodą LASSO pozwoliła ograniczyć liczbę zmiennych do 27. Przygotowanie danych obejmowało usunięcie zmiennych z ponad 10% brakujących wartości, inżynierię cech, kodowanie zmiennych binarnych oraz standaryzację zmiennych ciągłych przed zastosowaniem regresji logistycznej i LASSO. Postępowanie z brakującymi danymi zostało opisane rzetelnie, ponieważ przetestowano kilka metod imputacji, w tym MICE, imputację średnią, medianą i metodą k-najbliższych sąsiadów, a ostatecznie zdecydowano się na usuwanie listowe. Takie rozwiązanie zwiększa przejrzystość metodyczną, choć może ograniczać możliwość uogólniania wyników. Modele w publikacji (*Journal of Clinical Medicine* 2025; 14(11):3697) zbudowano z wykorzystaniem XGBoost, AdaBoost, LightGBM, lasu losowego oraz regresji logistycznej, co pozwoliło porównać metody złożone z modelem prostszym i bardziej interpretowalnym. Walidację wewnętrzną przeprowadzono za pomocą powtarzanej stratyfikowanej walidacji krzyżowej, a dostrajanie hiperparametrów wykonano przy użyciu biblioteki Hyperopt. Kolejna publikacja (*Journal of Clinical Medicine* 2025; 14(23):8325) została oparta na danych z Ogólnopolskiej Bazy Nadciśnienia Płucnego (BNP-PL), wieloośrodkowego, prospektywnego rejestru pacjentów z nadciśnieniem płucnym. Do analizy włączono 1755 dorosłych pacjentów z tętniczym nadciśnieniem płucnym, uwzględniając dane demograficzne, kliniczne, laboratoryjne, hemodynamiczne, echokardiograficzne, terapeutyczne oraz informacje o zdarzeniach śmiertelnych. Analizowano 165 zmiennych, a rozkład klas wynosił 1629:126. Zbiór podzielono na część uczącą i testową w proporcji 80:20, a selekcję cech przeprowadzono dwuetapowo: najpierw za pomocą LASSO, a następnie metodą sekwencyjnej eliminacji cech. Takie podejście należy ocenić pozytywnie, ponieważ ogranicza liczbę predyktorów, zwiększa przejrzystość modelu i może zmniejszać ryzyko nadmiernego dopasowania. Przygotowanie danych obejmowało usunięcie wartości odstających i duplikatów, obliczenie dodatkowych wskaźników klinicznych, takich jak wskaźnik objętości wyrzutowej oraz

stosunek TAPSE do skurczowego ciśnienia w tętnicy płucnej, a także kodowanie zmiennych kategoriowych metodą zero-jedynkową. W przypadku obserwacji zawierających braki danych zastosowano usuwanie listowe, uzasadniając ten wybór dążeniem do zachowania przejrzystości metodologicznej i ograniczenia ryzyka błędów związanych z imputacją. Modelowanie oparto na algorytmie XGBoost, wybranym ze względu na wysoką wydajność, szybkość obliczeń oraz możliwość wyjaśniania predykcji za pomocą wartości Shapleya. Proces uczenia przeprowadzono z użyciem powtarzanej stratyfikowanej walidacji krzyżowej, a dostrajanie hiperparametrów modelu XGBoost obejmowało m.in. maksymalną głębokość drzewa, tempo uczenia, liczbę estymatorów, parametry regularyzacji oraz parametry kontrolujące dobór cech i obserwacji w kolejnych iteracjach uczenia. Ze względu na silne niezrównoważenie klas przetestowano metody nadpróbkowania, takie jak SMOTE, ADASYN i SMOTETomek, jednak te nie przyniosły zadowalających rezultatów. Dobór metryk oceny modeli należy uznać za prawidłowy, ponieważ obejmował dokładność, swoistość, czułość, współczynnik korelacji Matthews, pole pod krzywą ROC oraz macierze pomyłek. Warto jednak rozważyć uzupełnienie raportowania wyników o precyzję, miarę F1 oraz pole pod krzywą precyzja-czułość. Bardzo dobrze należy ocenić wykorzystanie wartości Shapleya, ponieważ zwiększają one interpretowalność modeli i pozwalają określić znaczenie poszczególnych cech zarówno na poziomie populacji, jak i pojedynczego pacjenta. Całościowo metodyka jest nowoczesna, spójna i adekwatna do celu pracy. Jej mocną stroną jest wykorzystanie dwóch odmiennych źródeł danych klinicznych: dużej retrospektywnej bazy MIMIC-IV oraz prospektywnego, wieloośrodkowego rejestru BNP-PL. Istotną zaletą jest także świadome podejście do brakujących danych, niezrównoważenia klas, selekcji cech, walidacji krzyżowej, optymalizacji hiperparametrów i interpretowalności modeli. Opis środowiska obliczeniowego jest rzetelny, ponieważ zawiera wersję języka Python, zastosowane biblioteki oraz parametry sprzętu. W ostatecznej ocenie metodyka zasługuje na ocenę pozytywną, choć jej opis można by uzupełnić o bardziej szczegółowe przedstawienie przepływu danych, walidacji, kalibracji modeli oraz konsekwencji zastosowanego postępowania z brakującymi danymi.

Wyniki przedstawione w rozprawie należy uznać za spójne z celem pracy, wskazujące na możliwość zastosowania metod sztucznej inteligencji w predykcji ryzyka u pacjentów z chorobami sercowo-naczyniowymi. W publikacji (*Journal of Clinical Medicine* 2025; 14(11):3697) analiza statystyczna wykazała istotne różnice między pacjentami z migotaniem przedsionków leczonymi doustnymi antykoagulantami, którzy przeżyli, a tymi, którzy zmarli. Grupa zgonów charakteryzowała się starszym wiekiem, częstszym występowaniem niewydolności serca, demencji, paraplegii, nowotworów z przerzutami, ponownych przyjęć do szpitala, większym obciążeniem chorobami współistniejącymi oraz gorszymi wynikami laboratoryjnymi. Wyniki te są klinicznie wiarygodne, ponieważ wskazane czynniki odzwierciedlają cięższy stan ogólny

pacjentów, którzy zmarli. Wśród modeli predykcyjnych najlepszą skuteczność uzyskały algorytmy LightGBM i XGBoost, przy czym na zbiorze testowym LightGBM osiągnął AUC 0,886, dokładność 0,862, czułość 0,913, swoistość 0,859 oraz współczynnik korelacji Matthews'a 0,450. Szczególnie istotna jest wysoka czułość modelu, ponieważ w przewidywaniu śmiertelności ogranicza ryzyko pominięcia pacjentów rzeczywiście zagrożonych zgonem. Jednocześnie przewagę LightGBM nad XGBoost należy interpretować ostrożnie, ponieważ nie wykazano istotnych statystycznie różnic w czułości ani swoistości między tymi modelami. Na uznanie zasługuje wykorzystanie wartości Shapleya do interpretacji wyników. Największy wpływ na predykcję miały zmienne kliniczne, takie jak długość pobytu na oddziale intensywnej terapii, wskaźnik współchorobowości Charlsona, liczba współistniejących chorób i dni hospitalizacji oraz wybrane parametry laboratoryjne. Wskazuje to, że model nie tylko osiąga dobre parametry statystyczne, ale również opiera predykcje na zmiennych mających znaczenie kliniczne. W publikacji (*Journal of Clinical Medicine* 2025; 14(23):8325) wyniki są wartościowe, choć wymagają ostrożniejszej interpretacji. Pacjenci z tętniczym nadciśnieniem płucnym, którzy zmarli, byli starsi, częściej mieli przewlekłą chorobę nerek, cukrzycę, przebyty zawał mięśnia sercowego, hospitalizacje z powodu PAH, omdlenia, objawy niewydolności prawej komory serca, obecność płynu w worku osierdziowym, gorszy wynik 6-minutowego testu marszowego oraz wyższe stężenie NT-proBNP. Są to czynniki zgodne z obrazem bardziej zaawansowanej choroby i większego ryzyka klinicznego. Model XGBoost osiągnął umiarkowaną skuteczność: AUC 0,767, dokładność 0,738, czułość 0,800, swoistość 0,733 oraz współczynnik korelacji Matthews'a 0,298. Jego główną zaletą jest wysoka czułość, natomiast ograniczeniem pozostaje stosunkowo niski współczynnik korelacji Matthews'a oraz duża liczba wyników fałszywie dodatnich. Z tego względu model ten należy traktować jako potencjalne narzędzie wspomagające ocenę ryzyka, a nie jako samodzielną podstawę decyzji klinicznych. Interpretacja za pomocą wartości Shapleya wzmacnia wiarygodność uzyskanych wyników. Najważniejsze cechy wskazane przez model, takie jak stężenie NT-proBNP, dystans w 6-minutowym teście marszowym, stosunek TAPSE do sPAP oraz obecność płynu w worku osierdziowym, są zgodne z kliniczną oceną ryzyka pacjentów z PAH. Całościowo wyniki rozprawy są obiecujące i pokazują, że modele uczenia maszynowego mogą wspierać predykcję śmiertelności u pacjentów z migotaniem przedsionków oraz tętniczym nadciśnieniem płucnym. Najsilniejsze wyniki uzyskano w populacji pacjentów z migotaniem przedsionków, natomiast wyniki w populacji PAH są bardziej umiarkowane, ale nadal klinicznie interesujące. Mocną stroną pracy jest połączenie modelowania predykcyjnego z interpretacją za pomocą wartości Shapleya. Ograniczeniem pozostaje nie zrównoważenie klas, ryzyko wyników fałszywie dodatnich, brak szerzej omówionej kalibracji modeli oraz potrzeba dalszej walidacji w niezależnych populacjach. W ostatecznej ocenie wyniki należy uznać za dobrze udokumentowane, klinicznie uzasadnione i zgodne z aktualnym kierunkiem rozwoju predykcji klinicznej.

Dyskusję należy ocenić pozytywnie, ponieważ odnosi uzyskane wyniki do aktualnych problemów wdrażania sztucznej inteligencji w medycynie, w tym jakości danych, transparentności algorytmów oraz aspektów prawnych i etycznych. Autor trafnie zestawia wyniki własnego modelu dla pacjentów z migotaniem przedsionków z danymi z piśmiennictwa, wskazując zarówno przewagi opracowanego rozwiązania, jak i ograniczenia bezpośrednich porównań wynikające z różnych punktów końcowych oraz horyzontów predykcji. Mocną stroną dyskusji jest szczegółowe porównanie modelu z publikacji (*Journal of Clinical Medicine* 2025; 14(11):3697) z innymi modelami predykcyjnymi, w którym wykazano, że uzyskany wynik AUC 0,886 należy do lepszych rezultatów w analizowanej grupie badań. Równie wartościowa jest część dotycząca tętniczego nadciśnienia płucnego, ponieważ Doktorant odnosi wyniki własnego modelu do istniejących narzędzi prognostycznych i modeli sztucznej inteligencji stosowanych w PAH. Dyskusja właściwie podkreśla, że skuteczność modelu dla PAH jest porównywalna z częścią opisanych rozwiązań, ale niższa niż w najbardziej zaawansowanych modelach multimodalnych, co świadczy o wyważonej interpretacji wyników. Cennym elementem jest zwrócenie uwagi na potrzebę dalszej walidacji zewnętrznej, integracji danych z różnych źródeł oraz współpracy interdyscyplinarnej przed ewentualnym wdrożeniem modeli do praktyki klinicznej. Autor zasadnie wskazuje, że model dla AF osiągnął lepsze wyniki niż model dla PAH, jednocześnie wyjaśniając, że bezpośrednie porównanie obu modeli jest ograniczone przez różnice w populacjach, punktach końcowych, liczebności grup i czasie obserwacji. Dyskusja jest merytorycznie dojrzała i dobrze ugruntowana w literaturze, jednak miejscami wymaga korekty językowej oraz ostrożniejszego formułowania stwierdzeń dotyczących „wysokiej skuteczności”, zwłaszcza w odniesieniu do modelu PAH.

Na podstawie przeprowadzonych badań Autor sformułował następujące wnioski: 1) modele uczenia maszynowego, takie jak LightGBM i XGBoost, wykazują wysoką skuteczność w przewidywaniu śmiertelności pacjentów z chorobami sercowo-naczyniowymi, takimi jak AF i PAH; 2) w obu badaniach zastosowano metodę SHAP do wyjaśnienia predykcji modeli i oceny znaczenia poszczególnych cech. Interpretowalność wyników pozwala na identyfikację kluczowych czynników ryzyka dla indywidualnych pacjentów, co zwiększa praktyczną użyteczność modeli w klinice oraz pozwala lepiej zrozumieć na jakich cechach opiera się predykcja co zwiększa zaufanie do uzyskanych wyników; 3) modele osiągnęły wysokie wartości metryk, takich jak AUC, czułość i swoistość, co potwierdza ich skuteczność w przewidywaniu śmiertelności; 4) modele te mogą wspierać tradycyjne metody oceny ryzyka, oferując bardziej szczegółowe i spersonalizowane podejście do zarządzania pacjentami. Zastosowanie algorytmów uczenia maszynowego może przyczynić się do poprawy wyników leczenia i zmniejszenia obciążenia systemu opieki zdrowotnej w kontekście ekonomicznym oraz obciążenia pracą 5) wyniki badań wskazują na potencjał dalszego rozwoju

modeli predykcyjnych, w tym ich walidacji na innych populacjach i w różnych warunkach klinicznych. Sformułowane przez Doktoranta wnioski dobrze oddają uzyskane wyniki. Jednakże we wniosku 2 stwierdzenie, że modele w obu jednostkach chorobowych osiągnęły „wysoką skuteczność” nie jest w pełni uprawnione, ponieważ w PAH wyniki były raczej umiarkowane: AUC 0,767, czułość 0,800, swoistość 0,733 i MCC 0,298. Wniosek 4 także powinien być przerehabilitowany, ponieważ badania pokazują potencjał wspierania oceny ryzyka, ale nie wykazują bezpośrednio poprawy wyników leczenia, redukcji kosztów ani zmniejszenia obciążenia pracą. To są raczej możliwe skutki przyszłego wdrożenia, wymagające osobnej oceny klinicznej i ekonomicznej.

Reasumując należy stwierdzić, że cel pracy został w pełni zrealizowany, a uzyskane wyniki są oryginalnym osiągnięciem Autora, który dobrze opanował niełatwy warsztat badawczy i wykazał dobre merytoryczne przygotowanie. Uważam, że rozprawa doktorska spełnia warunki określone w art. 187 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce. Dlatego też zwracam się do Rady Dyscypliny Nauki Farmaceutyczne w Collegium Medicum im. L. Rydygiera w Bydgoszczy Uniwersytetu Mikołaja Kopernika w Toruniu o dopuszczenie mgr farm. Łukasza Ledzińskiego do dalszych etapów przewodu doktorskiego. Jednocześnie, ze względu na bardzo wysoki poziom merytoryczny pracy, docenionej również przez grono międzynarodowych ekspertów – recenzentów publikacji, wnioskuję o wyróżnienie pracy *summa cum laude*. Wyniki pracy opublikowano w renomowanych czasopismach o zasięgu międzynarodowym: *Journal of Clinical Medicine* (IF=2,9), *Journal of Clinical Medicine* (IF=2,9) i *Journal of Cardiovascular Development and Disease* (IF=2,3).



Prof. dr hab. n. med. Marek Drożdżik

