

Gliwice, 10.09.2026

Prof. Krzysztof Fajarewicz
Katedra Inżynierii i Biologii Systemów
Wydział Automatyki, Elektroniki i Informatyki
Politechnika Śląska

RECENZJA ROZPRAWY DOKTORSKIEJ

Nishadha Himanshi Hikkaduwa Liyanage

pt.

***„Advancing Explainability and Predictive Accuracy in AI-Driven Cardiovascular Risk
Assessment: A Hybrid Statistical and Neural Network Approach”***

przygotowanej na Wydziale Matematyki i Informatyki Uniwersytetu Łódzkiego

pod kierunkiem prof. Andrzeja Nowakowskiego

oraz dr Marty Lipnickiej

1. Przedmiot i zakres rozprawy

Przedłożona do recenzji rozprawa doktorska dotyczy zastosowania metod sztucznej inteligencji, uczenia maszynowego i analizy przeżycia do oceny ryzyka chorób sercowo-naczyniowych. Tematyka pracy jest aktualna i ma istotne znaczenie zarówno z punktu widzenia informatyki, jak i potencjalnych zastosowań biomedycznych. Szczególnie ważnym zagadnieniem, podejmowanym przez Autorkę, jest próba pogodzenia wysokiej jakości predykcji z interpretowalnością modelu, co pozostaje jednym z zasadniczych problemów współczesnych zastosowań metod AI w medycynie.

W rozprawie rozważane są dwa podejścia. Pierwsze stanowi klasyfikator oparty na wielowarstwowej sieci neuronowej typu Feedforward Neural Network (FFNN). Drugie, przedstawiane jako główne autorskie rozwiązanie, integruje nienadzorowaną klasteryzację Winner-Takes-All (WTA), modele proporcjonalnego hazardu Coxa budowane osobno dla

zidentyfikowanych podgrup pacjentów oraz miarę Cumulative Prevalence Ratio (CPR), służącą do przypisania klastrom skumulowanego poziomu ryzyka.

Analizy przeprowadzono na zbiorze CVD SL obejmującym 66 816 obserwacji pacjentów. Zbiór zawiera dziewięć zmiennych demograficznych, klinicznych i związanych ze stylem życia: wiek, płeć, BMI, rozkurczowe ciśnienie tętnicze, poziom cholesterolu, palenie tytoniu, spożywanie alkoholu, aktywność fizyczną oraz stężenie glukozy na czczo. Zmienną wynikową jest binarna informacja o obecności choroby sercowo-naczyniowej.

Zakres tematyczny rozprawy należy zatem uznać za właściwy dla dyscypliny informatyka, a podjęty problem badawczy za istotny i aktualny.

2. Struktura i ogólna ocena rozprawy

Rozprawa ma przejrzystą strukturę. Po części wprowadzającej i analizie eksploracyjnej danych Autorka przedstawia podstawy sztucznych sieci neuronowych i wyniki eksperymentów z FFNN, a następnie szczegółowo opisuje proponowany framework hybrydowy. Końcowa część zawiera dyskusję wyników, porównanie obu podejść oraz wskazanie potencjalnych kierunków dalszych badań.

Do zalet pracy zaliczam znaczną liczebność analizowanego zbioru, próbę połączenia metod statystycznych z algorytmami uczenia maszynowego oraz udostępnienie kodu implementującego istotne części eksperymentów. Ułatwia to ocenę rzeczywistego przebiegu analiz i zwiększa możliwość ich reprodukcji. Wartościowym elementem jest również próba modelowania heterogeniczności populacji pacjentów poprzez tworzenie podgrup zamiast stosowania wyłącznie jednego globalnego modelu.

Jednocześnie rozprawa zawiera szereg problemów metodologicznych, które w mojej ocenie mają znaczenie zasadnicze dla interpretacji przedstawionych wyników.

3. Ocena części dotyczącej FFNN

W części poświęconej sieci neuronowej Autorka testuje kilka architektur FFNN, spośród których wybiera model o strukturze 9–10–20–10–1. Sieć służy do bezpośredniej klasyfikacji binarnej CVD i wykorzystuje standardową funkcję sigmoid na wyjściu.

Sam eksperyment jest poprawnie sformułowany jako problem klasyfikacyjny, jednak sposób wyboru architektury budzi zastrzeżenia. Poszczególne warianty sieci są oceniane na zbiorze testowym, a następnie spośród nich wybierana jest architektura o najlepszym wyniku. Oznacza to, że zbiór testowy zostaje wykorzystany do selekcji modelu, a raportowany wynik przestaje być w pełni niezależną oceną zdolności generalizacyjnej modelu.

Dziwnie wygląda kształt krzywej ROC na Rysunku 2.7. Składa się on z dwóch odcinków. Co więcej, jest on inny niż na Rysunku 4.1, gdzie ta sama krzywa ROC porównana jest z krzywą ROC dla proponowanego podejścia hybrydowego.

Brakuje również systematycznego porównania FFNN z innymi klasycznymi metodami klasyfikacyjnymi na dokładnie tych samych danych i według identycznego protokołu walidacji. Dla zbioru zawierającego jedynie dziewięć cech, przy bardzo dużej liczbie obserwacji, naturalnymi punktami odniesienia powinny być na przykład regresja logistyczna, SVM, Random Forest, kNN.

Znaczące jest przy tym, że w kodzie analiz hybrydowych znajduje się odrębna regresja logistyczna, która dla tych samych dziewięciu cech osiąga AUC około 0,868. Wynik ten sugeruje, że prosty model statystyczny może osiągać jakość predykcji wyraźnie wyższą niż raportowana dla FFNN. Regresja logistyczna nie została jednak potraktowana jako pełnoprawny benchmark dla FFNN.

4. Ocena proponowanego frameworku hybrydowego

Główna propozycja Autorki ma strukturę:

$$WTA \rightarrow \text{cluster-specific Cox models} \rightarrow CPR \rightarrow \text{risk category}.$$

Autorka rozpoczyna od 15 prototypów WTA. Po zakończeniu uczenia aktywnych pozostaje osiem klastrów, z których jeden zawiera zaledwie jednego pacjenta. Dla tego klastra model Coxa nie jest estymowany, wobec czego dalsze modelowanie obejmuje siedem klastrów.

Koncepcja budowania osobnych modeli dla podgrup pacjentów jest interesująca i potencjalnie pozwala na uwzględnienie heterogeniczności populacji. Niemniej sposób konstrukcji i walidacji rozwiązania nasuwa kilka istotnych pytań.

Po pierwsze, w rozprawie nie przedstawiono kompletnej reguły predykcji dla przypadku, gdy nowy pacjent zostanie przypisany do prototypu, dla którego nie istnieje model Coxa. Funkcja WTA może wskazać dowolny wyuczony prototyp, podczas gdy część survivalowa istnieje jedynie dla siedmiu klastrów. Nie określono, czy taki pacjent powinien być przypisany do najbliższego poprawnego klastra, czy też zastosowany powinien być inny mechanizm zastępczy.

Po drugie, zasadnicze zastrzeżenie budzi sama konstrukcja analizy przeżycia. Dane mają charakter przekrojowy, natomiast Autorka jako zmienną czasu do wystąpienia choroby wykorzystuje wiek pacjenta, a jako wskaźnik zdarzenia – obecność CVD. W rozprawie przyznano, że zastosowanie wieku jako zmiennej survivalowej wynika z braku rzeczywistych danych longitudinalnych. W praktyce oznacza to jednak, że pacjent w wieku 60 lat z CVD jest traktowany jak osoba, u której zdarzenie wystąpiło w wieku 60 lat, choć choroba mogła zostać rozpoznana znacznie wcześniej.

Taka reprezentacja danych nie odpowiada klasycznemu schematowi danych time-to-event z prawostronnym cenzorowaniem i wymagałaby bardzo silnego uzasadnienia. Dodatkowo wiek jest jednocześnie cechą wykorzystywaną przy tworzeniu klastrów. Tym samym zmienna uczestnicząca w konstrukcji podgrup staje się następnie osią czasu w modelu przeżycia. Może to powodować, że część różnic między funkcjami przeżycia klastrów jest generowana przez sam sposób ich konstrukcji.

5. Miara CPR i sposób definiowania ryzyka

W rozprawie dla klastra k definiowane jest Cumulative Prevalance Ratio (CPR) – wzór 3.9.

Następnie wartości CPR są dzielone na cztery kategorie ryzyka: Low, Moderate, High i Critical, z wykorzystaniem progów odpowiadających wybranym kwantylom rozkładu CPR.

Należy podkreślić, że CPR jest wielkością na poziomie klastra. Ponieważ modelowana jest wartość dla średniego profilu, każdy pacjent przypisany do tego samego klastra otrzymuje identyczne CPR. W wyniku pozostawienia siedmiu klastrów istnieje zaledwie siedem możliwych wartości CPR, które następnie redukowane są do czterech kategorii ryzyka.

Ogranicza to w istotny sposób deklarowaną personalizację modelu. Warto zauważyć, że sam model Coxa pozwalałby na obliczanie indywidualnego liniowego predyktora dla każdego

pacjenta, natomiast zastosowana definicja CPR eliminuje tę indywidualną zmienność po przypisaniu pacjenta do klastra.

6. Uwagi dotyczące zbioru danych CVD SL

Ponieważ zbiory użyte w Rozprawie i we wcześniejszym artykule Doktorantki [Liyange & Lipnicka, 2026] mają taką samą liczbę obserwacji i taką samą nazwę zakładam, że to te same zbiory. Tabela 1.1 w Rozprawie i Tabela 1 w artykule podają część statystyk opisowych danych. Dziwne jest to, że się różnią. Dotyczy to szczególnie zakresu zmiennych Min-Max. Dziwnie wyglądają również histogramy dla wieku (Rysunek 1.1) i BMI (Rysunek 1.3). Histogram wieku zawiera dziwną składową okresową o okresie 2 lata, a histogram dla BMI nie przypomina typowego rozkładu log-normalnego obserwowanego zazwyczaj w populacjach ludzkich.

7. Najpoważniejsza wątpliwość dotycząca AUC modelu hybrydowego

Za najistotniejszy problem empiryczny rozprawy uważam sposób raportowania wartości $AUC=0,8676$ dla modelu hybrydowego.

Na rysunku 4.1 przedstawiono gładką krzywą ROC przypisaną „Hybrid Framework”. Tego rodzaju krzywa wymaga ciągłego lub przynajmniej wielopoziomowego indywidualnego score’u. Tymczasem CPR ma tylko siedem możliwych wartości, a po kategoryzacji jedynie cztery poziomy. ROC utworzony bezpośrednio z CPR powinien zatem mieć bardzo ograniczoną liczbę punktów operacyjnych i wyraźnie schodkowy charakter.

Jednocześnie w załączonym kodzie znajduje się odrębny eksperyment z regresją logistyczną, wykorzystujący dziewięć pierwotnych cech wejściowych.

Jeżeli $AUC=0,8676$ rzeczywiście pochodzi z pomocniczej regresji logistycznej, to wynik ten nie może być traktowany jako miara jakości WTA–Cox–CPR. Problem ten wymaga jednoznacznego wyjaśnienia poprzez wskazanie konkretnej zmiennej predykcyjnej wykorzystanej do wygenerowania ROC, sposobu jej obliczenia oraz procedury walidacji.

Podobna niejasność dotyczy raportowanej accuracy modelu hybrydowego. CPR generuje cztery kategorie ryzyka, podczas gdy accuracy odnosi się do przewidywania klasy. W rozprawie nie przedstawiono w sposób wystarczająco jednoznaczny reguły zamieniającej wynik frameworku na binarną decyzję $CVD=0/1$.

8. Interpretowalność i wyjaśnialność

Pojęcie „explainability” jest silnie eksponowane w tytule i celach rozprawy. W przypadku FFNN nie przeprowadzono jednak faktycznej analizy XAI. Autorka wprost stwierdza, że wagi sieci nie są interpretowane. Nie zastosowano SHAP, permutation importance, analizy wrażliwości, metod ablacyjnych ani innych typowych narzędzi pozwalających określić wpływ poszczególnych zmiennych na predykcję sieci.

W przypadku modeli Coxa istnieje naturalny mechanizm interpretacji poprzez współczynniki β i hazard ratios. Potencjalnie byłyby to jeden z najmocniejszych elementów proponowanego podejścia. W samej rozprawie brak jednak systematycznej analizy pokazującej, które czynniki są najistotniejsze w poszczególnych klastrach i jak zmieniają się ich efekty między podgrupami.

Co znamienne, materiał tego rodzaju znajduje się we wcześniejszym artykule Liyanage i Lipnickiej, gdzie zaprezentowano m.in. współczynniki Coxa, błędy standardowe, p-values, hazard ratios i przedziały ufności oraz rankingi istotności cech.

Nie udało mi się również znaleźć rzeczywistego etapu selekcji cech. Wszystkie dziewięć cech jest używanych w FFNN i WTA, a wiek zostaje pominięty w modelu Coxa ze względu na pełnienie funkcji zmiennej czasu, a nie w wyniku procedury selekcji cech.

9. Oryginalność pracy i relacja do wcześniejszej publikacji

Kluczowe znaczenie dla oceny oryginalności ma publikacja H. Liyanage i M. Lipnickiej z 2026 r. pt. Computational framework for dynamic cardiovascular risk assessment with cluster-specific Cox models and cumulative risk analysis. Artykuł przedstawia już zasadniczą koncepcję integrującą competitive learning, klasteryzację pacjentów, osobne modele Coxa i CPR. CPR jest w nim przedstawiane jako kluczowa nowość.

Sama rozprawa jednoznacznie stwierdza, że proponowany framework bezpośrednio bazuje na tej publikacji i że wcześniejsza praca stanowi jego metodologiczną podstawę. Jako rozszerzenia wskazywane są przede wszystkim pełniejsza formalizacja, szersza walidacja i bardziej szczegółowa analiza WTA.

Nie jest to samo w sobie nieprawidłowością. Współautorskie publikacje mogą stanowić podstawę rozprawy doktorskiej. Konieczne jest jednak precyzyjne rozgraniczenie wkładu

indywidualnego doktorantki od wkładu współautorki. W rozprawie nie znalazłem jednoznacznego opisu indywidualnego wkładu Autorów wcześniejszej publikacji.

Dodatkowo definicja CPR w artykule i w rozprawie nie jest identyczna. W artykule CPR jest określone jako całość ze skumulowanego hazardu po przedziale wieku 20–99, podczas gdy w rozprawie jest to przyrost skumulowanego hazardu bazowego przemnożony przez czynnik wynikający z liniowego predyktora. Zmiana definicji centralnej, autorskiej miary powinna zostać wyraźnie wyjaśniona i uzasadniona.

9. Uwagi formalne

Poziom językowy rozprawy jest zasadniczo komunikatywny, jednak tekst wymagałby staranniejszej redakcji. Występują nierozwiązane odsyłacze, np. „see Section ??” - strona 62.

Styl rozprawy jest miejscami nadmiernie deklaratywny. Sformułowania dotyczące „strong explainability”, „clinical relevance”, „personalised prediction” czy „generalisation” są używane częściej, niż wynikałoby to z zakresu przedstawionych analiz empirycznych.

10. Pytania do Autorki rozprawy

1. Jaka dokładnie zmienna predykcyjna została wykorzystana do skonstruowania krzywej ROC przedstawionej na rysunku 4.1 i uzyskania $AUC=0,8676$? Czy był to CPR, liniowy predyktor Coxa, czy prawdopodobieństwo z pomocniczej regresji logistycznej?
2. W jaki sposób uzyskano raportowaną accuracy modelu hybrydowego, skoro ostatecznym wynikiem CPR są cztery kategorie ryzyka, a nie binarna predykcja CVD?
3. Jakie jest metodologiczne uzasadnienie traktowania aktualnego wieku pacjenta jako czasu do zdarzenia w przekrojowym zbiorze danych i w jaki sposób należy interpretować cenzorowanie w tym schemacie?
4. Dlaczego w Rozprawie nie podjęto próby selekcji cech?
5. Jaki był indywidualny wkład doktorantki w artykuł Liyanage i Lipnickiej z 2026 r. i które elementy frameworku przedstawionego w rozprawie stanowią nowe osiągnięcie w stosunku do tej publikacji?
6. Dlaczego definicja CPR zastosowana w rozprawie różni się od definicji opublikowanej w artykule z 2026 r. i jakie są konsekwencje tej zmiany?

11. Konkluzja

Rozprawa dotyczy aktualnego i ważnego problemu badawczego, zawiera interesującą koncepcję modelowania ryzyka w podgrupach pacjentów oraz prezentuje obszerny materiał eksperymentalny. Za pozytywne elementy pracy uznaję przede wszystkim próbę połączenia metod uczenia maszynowego z analizą przeżycia, wykorzystanie dużego zbioru danych oraz dążenie do zwiększenia interpretowalności modeli stosowanych w medycynie.

Jednocześnie stwierdzone przeze mnie problemy nie mają wyłącznie charakteru redakcyjnego. Część z nich dotyczy centralnych założeń metodologicznych oraz sposobu walidacji głównej propozycji badawczej. Szczególnie poważne są wątpliwości związane z zastosowaniem analizy przeżycia do danych przekrojowych i sposobem wykorzystania wieku jako czasu do wystąpienia choroby, niejednoznaczny charakter CPR oraz przede wszystkim pochodzeniem raportowanego $AUC=0,8676$ przypisanego frameworkowi hybrydowemu.

W recenzji, oprócz wskazania zalet badanego modelu i uzyskanych wyników, zamieściłem szereg uwag krytycznych oraz aspektów wymagających lub wartych wyjaśnienia. Jestem jednak przekonany, że w trakcie obrony Doktorantce uda się odpowiedzieć na wątpliwości zawarte w niniejszej recenzji.

Stwierdzam zatem, że rozprawa, którą recenzuję, spełnia w mojej ocenie standardy stawiane rozprawom doktorskim, zarówno ustawowe (to jest zgodne z art. 187 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce, z późniejszymi zmianami) jak i te zwyczajowe. Wnoszę zatem o dopuszczenie Pani Nishadha Himanshi Hikkaduwa Liyanage do dalszych etapów postępowania o nadanie stopnia doktora.

.....