

Ocena

Rozprawy doktorskiej Lek. Anny Poradzkiej p.t.

Wykorzystanie sztucznej sieci neuronowej jako modelu predykcyjnego sześciomiesięcznego rokowania w zespole stopy cukrzycowej

Pomimo postępu w terapii cukrzycy oraz systematycznego wprowadzania nowych leków, narzędzi do monitorowania glikemii i wiedzy o patomechanizmach powikłań, zespół stopy cukrzycowej pozostaje istotnym problemem klinicznym. Dane epidemiologiczne wskazują, że jeden na siedmiu chorych na cukrzycę zmierzy się z tym powikłaniem w ciągu swojego życia. Konsekwencje zespołu obejmują wyłączenie z życia zawodowego, pogorszenie jakości życia, a w przypadkach wymagających amputacji skutkują ciężką niepełnosprawnością. Z tego względu dostęp do metod pozwalających efektywnie diagnozować, leczyć i przewidywać efekty kliniczne terapii w zespole stopy cukrzycowej są kluczowym dla diabetologii obszarem badań. W obszarze diagnostyki i oceny klinicznych efektów leczenia, szczególnie dynamicznie rozwijającym się obszarem badań jest wykorzystanie zaawansowanych narzędzi statystycznych, w tym wykorzystujących narzędzia oparte o sztuczną inteligencję i uczenie maszynowe. Zastosowanie tych skomplikowanych technik do stworzenia efektywnego systemu predykcyjnego względem leczenia zespołu stopy cukrzycowej stanowi oś pracy doktorskiej lek. Anny Poradzkiej.

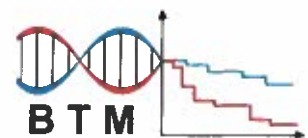
Szczegółowe cele pracy zdefiniowano w sposób poprawny, typowy dla tego typu projektów. Ze względu na złożoność i kompleksowość projektu, nie dziwi ich znacznie większa niż typowo liczba. Wymienione cele składają się w sposób logiczny i uporządkowany w spójną całość. Należy podkreślić, że są to cele bardzo wartościowe, o istotnym elemencie klinicznym i praktycznym co zasługuje na uznanie.

Oceniana praca ma układ typowy dla prac klinicznych. Po obszernym wstępie, przedstawiona jest rozbudowana sekcja metody omawiająca zastosowane narzędzia badawcze, a następnie obszerna sekcja wyników, po czym następuje ich omówienie w dyskusji i podsumowanie. Główny tekst pracy doktorskiej liczy 208 stron maszynopisu. Na uwagę zasługuje bogata oprawa graficzna pracy, przedstawiona na 42 rycinach oraz obszerne przedstawienie danych w 102 tabelach.

We wstępie pracy Autorka koncentruje się na opisie epidemiologii cukrzycy, jej powikłań oraz efektywności leczenia stopy cukrzycowej w Polsce i na Świecie. Rozdział ten jest dobrze napisany i dotyka zarówno aspektów zachorowalności, chorobowości jak i rokowania pacjentów u których to

Department of Biostatistics
& Translational Medicine

92-215 Łódź, 15 Mazowiecka st.
e-mail: btm @umed.lodz.pl
tel. (042) 272 53 85
www.biostat.umed.pl



powikłanie cukrzycy się rozwija. Autorka słusznie zwraca uwagę na aspekt jakości życia i pozamedyczne, społeczne konsekwencje rozpoznania i leczenia stopy cukrzycowej, co wskazuje na jej kompleksowe i wielopłaszczyznowe podejście do badanej problematyki. We wstępie przytoczone są aktualne dane naukowe i epidemiologiczne, ukazany jest kontekst kliniczny oraz efektywność rutynowo stosowanych technik leczenia. Autorka przytacza również dane NFZ wykazujące dokładne liczebności oraz zagregowane cechy demograficzne pacjentów leczonych z powodu stopy cukrzycowej w Polsce oraz podkreśla ekonomiczne korzyści związane ze skuteczną prewencją tego powikłania. Kolejne sekcje wstępu objaśniają rolę różnych technik diagnostycznych oceniających funkcję naczyń i nerwów których uszkodzenie odgrywa ważną rolę w patogenezie i procesie leczenia. W sposób zwięzły i uwzględniający aktualne piśmiennictwo ukazane są zalety, wady i wartości referencyjne kluczowych badań diagnostycznych. Sekcja ta uzasadnia również wykorzystane w projekcie metody na wynikach których Autorka opierała swoje modele klasyfikacyjne. Miejscami, wstęp jest nawet zbyt obszerny i porównania efektywności diagnostycznej metod tamże omawianych mogłyby równie dobrze zostać przeniesione do dyskusji, gdzie efektywność tę można odnieść do uzyskanego przez Autorkę modelu predykcyjnego. Niemniej, kompleksowe oraz wnikliwe zebranie i podsumowanie wielu lat badań nad przedmiotowym powikłaniem cukrzycy, świadczy o dojrzałości Autorki jako pracownika naukowego i jej dogłębnej znajomości tematu.

Metody zastosowane w pracy opisano ze szczegółami technicznymi na 28 stronach rozprawy. Sekcję otwiera wyliczenie założeń pracy, które jest miejscami dość chaotyczne – zestawia bowiem konkretne problemy kliniczne z omówieniem i dywagacjami na temat przyczyn niektórych z nich. Następna sekcja definiuje cele pracy które są logicznie sformułowane i jasno tłumaczą zakres dalszych działań badawczych opisanych w pracy.

Badanie zostało przeprowadzone w Klinice Diabetologii i Chorób Wewnętrznych Centralnego Szpitala Klinicznego Uniwersyteckiego Centrum Klinicznego Warszawskiego Uniwersytetu Medycznego po zasięgnięciu opinii Komisji Bioetycznej WUM. Rekrutacja pacjentów trwała rok i szczęśliwie została zakończona przed wybuchem pandemii SARS-CoV2.

Liczebność grupy została oceniona na podstawie dostępności pacjentów i post-factum weryfikowana regułą Kalayeh co do możliwości przeuczenia sieci neuronowej. Jest to strategia wykorzystywana w tego typu badaniach, jakkolwiek można było liczebność próby oszacować na podstawie porównania efektywności klasyfikacyjnej względem innych technik. Ryzyko podjęte przez Doktorantkę przy przyjętej strategii było dość duże i mogło skutkować niemożnością wytworzenia dobrego modelu bez jego przeuczania.

Kryteria włączenia i wyłączenia pacjentów były jasno zdefiniowane, procedura włączania pacjentów do badania obserwacyjnego nie budziła wątpliwości. Dla czystego formalizmu sprawozdawczego, w sekcji wyniki, przydałby się w pracy wykres typu flowchart obrazujący liczbę pacjentów w ogóle, pacjentów spełniających kryteria włączenia, wyłączonych, utraconych i dostępnych do oceny w obserwacji odległej.

Metody zastosowane w ocenie klinicznej funkcji układu krążenia u pacjentów są zasadne, sposób gromadzenia i przetwarzania danych nie budzi zastrzeżeń. Fakt utraty możliwości pomiaru TcPO₂ w podgrupie pacjentów skomplikował plan badania i pewnym niedociągnięciem jest brak zabezpieczenia się Badaczki na wypadek takiej ewentualności. Podobnie techniki oceniające obecność i stopień nasilenia neuropatii, zakres owrzodzenia, cechy radiologiczne i pozostałe aspekty zespołu stopy cukrzycowej zostały wybrane odpowiednio i wdrożone w należyty sposób. Ocenę efektów leczenia przeprowadzano po 3 i 6 miesiącach od włączenia do badania. Kolejnym pozytywnym aspektem pracy było wykorzystanie technik telemedycznych do oceny punktów końcowych u pacjentów niemogących stawić się osobiście na wizycie ze względu na trudności w transporcie lub ogólnie zły stan zdrowia. Autorka opisuje logistykę tego procesu wskazując na trudności związane z przedłużoną opieką nad pacjentami o ograniczonej możliwości poruszania się i wykazuje dokładną znajomość procesu, co potwierdza jej znaczne zaangażowanie w projekt i determinację do jego poprawnej realizacji.

Opis metod statystycznych jest częścią szczególnie interesującą w tej pracy. Opis działania sieci jest miejscami zbyt rozległy i mógłby zostać bez szkody dla pracy przeniesiony do wstępu lub usunięty całkowicie celem skrócenia rozprawy (choćby powtórzenie reguły Kalayeh). Niemniej, sekcja ta jest interesująca i ukazuje znajomość technik stosowanych w analizie danych przez Autorkę.

Sekcja wyników badań, została przedstawiona na siedemdziesięciu dwóch stronach maszynopisu w bardzo przejrzysty sposób. Wyniki wykazują olbrzymi nakład pracy Doktorantki. Efekty leczenia po 3 lub 6 miesiącach nie odbiegały znacząco od oczekiwań literaturowych. Selekcja wstępna zmiennych przeprowadzona przez Autorkę opierała się na porównaniach grup w zależności od efektów leczenia i jest metodą rutynowo wykorzystywaną w takich zadaniach. Podjęto także próbę wykorzystania liniowej analizy dyskryminacyjnej. Dla czystości „walidacji” modelu, proponowałbym natomiast przeprowadzać taką selekcję zmiennych w oparciu o przypadki przydzielone jedynie do próby uczącej lub kalibracyjnej (lub testowej, wykorzystywanej do optymalizacji hiperparametrów modelu). Czyszczenie danych w oparciu o próbki z próby walidacyjnej ułatwia późniejszą konstrukcję modelu, ale naraża na dyskusję z recenzentami co do niezależności grupy walidacyjnej. Ostatecznie, po podziale na grupy Autorka dobrze charakteryzuje efektywność procedury przydziału i przystępuje do konstrukcji sieci neuronowych. Wybrana liczba 500 sieci tworzonych w celu predykcji efektów leczenia po 3 miesiącach była dość mała z punktu widzenia liczby możliwych kombinacji parametrów (liczba neuronów * funkcje aktywacji 1 * funkcje aktywacji 2), jednak cele analityczne zostały mimo to osiągnięte w postaci sieci o dokładności ~80% na zbiorze walidacyjnym. Z punktu widzenia przejrzystości wyników proponowałbym w tym miejscu prezentować krzywą ROC dla zbioru uczącego (i ewentualnie testowego) oraz macierz klasyfikacji dla zbioru walidacyjnego oddzielnie. W przedstawionej tabeli 9.1.16 i podobnych trudno jest odgadnąć który ze zbiorów ważył najbardziej pod względem błędnych klasyfikacji. Jednocześnie, mając pulę dobrze działających sieci, warto było dokładniej przyrzeć się źle klasyfikowanym przypadkom – czy były to te same czy każdorazowo inne osoby? Takie spostrzeżenie mogłoby nakierować Autorkę na trop modeli złożonych z różnych sieci lub

metod klasyfikacyjnych adekwatnych dla przypadków typowych lub unikalnych. Przeprowadzona analiza wrażliwości pozwoliła na usunięcie nadmiarowych zmiennych i doprowadzenie do modelu klasyfikacyjnego o możliwie najmniejszej złożoności, adekwatnej do problemu badawczego za pomocą automatycznego kreatora sieci. Następnie, Autorka podjęła próbę konstrukcji lepszej sieci w oparciu o ręczną selekcję funkcji aktywacji i liczby neuronów warstwy ukrytej. Proces ten był dużo bardziej żmudny i de facto replikował proces automatyczny. Efektem były sieci o efektywności zbliżonej do najlepszych sieci wytworzonych poprzez metodę grid search z nieco większą preferencją co do swoistości kosztem czułości (tabela 9.1.46). Zarówno sieci automatycznie jak i ręcznie wytworzone wykazywały przewagę nad modelem prostej regresji logistycznej potwierdzając przypuszczenia Autorki co do potencjału ANN w tym problemie badawczym.

Wyniki analiz co do rokowania w horyzoncie 6 miesięcy wskazywały dość zaskakująco wyższą trafność automatycznie wygenerowanych sieci w zbiorze walidacyjnym niż pozostałych, co było prawdopodobnie efektem „szczęśliwego” losowania przypadków. Ostatecznie, po oczyszczeniu modelu ze zmiennych redundantnych dokładność w zbiorze walidacyjnym osiągnęła bardzo przyzwoite 86% wskazując na możliwą użyteczność kliniczną takiego narzędzia. Ręcznie projektowane przez Autorkę sieci zbliżyły się do tego wyniku i, podobnie jak w przypadku obserwacji 3-miesięcznej, przewyższały trafność modelu regresji logistycznej. Łącznie sekcja wyników, ilustrowana za pomocą kilkudziesięciu klarownych tabel i rycin w czytelny i logiczny sposób ilustruje plan i wykonanie badań zmierzających do realizacji zamierzonych celów pracy. Ostatnią sekcją wyników jest opis wytworzonego narzędzia – kalkulatora prawdopodobieństwa efektu leczenia w oparciu o kilka kluczowych zmiennych. Narzędzie to, możliwe do wdrożenia poprzez narzędzie przeliczające dane w oparciu o utworzony plik pmml stanowi oryginalny i cenny wynik projektu możliwy do zastosowania w praktyce klinicznej i dalszych projektach naukowych.

Dyskusja pracy została przeprowadzona w sposób klasyczny. Autorka omawia główne wyniki pracy, zwracając uwagę na kliniczne i psychospołeczne aspekty leczenia pacjenta z zespołem stopy cukrzycowej. Nakreśla scenariusze w jakich możnaby stosować utworzony model ANN. Sekcja dotycząca zastosowanych modeli mogłaby być uzupełniona o bardziej złożone modele (sieci głębokie, sieci konwolucyjne, sieci z autoenkoderami, itp.) aczkolwiek uwzględniając, że praca dotyczy dyscypliny nauk medycznych, skłaniałbym się raczej ku minimalizacji tego aspektu dyskusji niż jej ekspansją w kierunku nauk ścisłych. Autorka omawia problemy i rozwiązania jakie napotkała oraz identyfikuje „słabe” strony pracy, przy czym dotyczą one tak naprawdę mankamentów metod jako takich (losowy podział, restrykcje treningu sieci itp.) a nie błędnego podejścia. Kolejne sekcje dyskusji uwypuklają znaczenie aspektów metodycznych i klinicznych i ich możliwe oddziaływanie na efektywność tworzonego modelu klasyfikacyjnego. Każdy podrozdział jest omówiony zwięźle w kontekście bieżącego piśmiennictwa, co pozwala ocenić jakość i znaczenie wyników uzyskanych przez Doktorantkę w kontekście aktualnego stanu wiedzy.

Piśmiennictwo obejmuje aż 277 pozycji, w olbrzymiej większości opublikowane w ostatnich latach w czasopiśmie z Impact Factor, co świadczy o aktualności poruszanego tematu i olbrzymim

zainteresowaniu tym obszarem badań biomedycznych. Treść pracy zamyka zwięzła sekcja wnioski, w syntetyczny i logiczny sposób formułująca wynikające z wyników pracy wnioski dobrze podsumowujące znaczenie i implikacje przedstawionych w pracy odkryć. Autorka, w rozprawie zamieszcza dodatkowo aneks z materiałami graficznymi poddanymi analizie („Zdjęcia owrzodzeń pacjentów objętych badaniem”) oraz potwierdzenie akceptacji realizacji projektu badawczego przez Komisję Bioetyczną przy Warszawskim Uniwersytecie Medycznym.

Ogółem pod względem merytorycznej zawartości, poprawności zaplanowania badań, ich przeprowadzenia, jak również pod względem ogólnej formy i organizacji treści pracy, rozprawę lek. Anny Poradckiej oceniam bardzo wysoko. Na szczególne podkreślenie zasługuje bardzo kompleksowe podejście do analizy statystycznej danych. Procedura oczyszczenia zbioru danych z artefaktów, identyfikacja zmiennych współliniowych, konstrukcja modeli klasyfikacyjnych i ich dalsza obróbka poprzez analizę wrażliwości i dążenie do wytworzenia narzędzia maksymalnie użytecznego klinicznie świadczą o wysokiej świadomości Autorki co do możliwości metod statystycznych oraz potwierdzają jej wiedzę i umiejętności w zakresie umiejętnego stosowania szerokiego wachlarza narzędzi statystycznych i data miningowych.

Chociaż w całości rozprawa nie budzi wątpliwości w zakresie oryginalności tematyki badawczej, zastosowanej metodyki badań, czy interpretacji wyników tych badań, po jej lekturze nasunęło mi się kilka pytań, na które odpowiedzi nie byłem w stanie wydedukować w oparciu o tekst pracy. Część z nich wynika z mojej własnej badawczej ciekawości i, jakkolwiek zapewne przekracza zakres wykonanych lub planowanych prac, być może stanie się przyczynkiem do ciekawej dyskusji lub kolejnych projektów:

- Rozpoczynając pracę nad systemem sztucznej inteligencji w prognozowaniu rokowania możnaby pokusić się o stworzenie narzędzia bazującego również na zdjęciach owrzodzeń i ekstrakcji cech obrazu odzwierciedlających lokalizację, ciężkość i zakres procesu chorobowego. Czy Autorka rozważyła „pójście dalej” w tym kierunku i eskalację wykorzystania technik uczenia maszynowego ponad analizę danych zbieranych ręcznie, w kierunku analizy obrazu?
- Czy omówione w sekcji 10.3.2 problemy związane z włączeniem pacjentów z owrzodzeniami obu stóp nie byłyby rozwiązywalne poprzez dodanie do analizy zmiennej „chora druga stopa”?
- Czy częstsza ocena kliniczna nie pozwoliłaby na udoskonalenie modelu i możliwość predykcji czasu do zagojenia a nie jedynie punktu końcowego w dwóch interwałach? Pozwoliłoby to prawdopodobnie na mniejszy dropout pacjentów i ocenę dynamiki gojenia się/pogorszenia w zależności od predykcji ANN w czasie.
- Skoro Autorka sugeruje porównanie efektywności wytworzonej sieci ze skalą PEDIS, co stało na przeszkodzie by zrobić to w zebranych materiale?

Na uznanie zasługuje dbałość o styl i formę pracy, w której, mimo obszernego tekstu trudno dopatrzeć się zwyczajowych błędów literowych lub stylistycznych. Niektóre zdania mają co prawda dość dziwną strukturę logiczną lub stylistyczną (np. „Cukrzyca jest ściśle powiązana zarówno z powikłaniami mikro – jak i makro-angiopatycznymi”; „Na wskaźnik paluch-ramię wpływ wywiera

również choroba Raynauld i twardzina dlatego wskaźnik zawodzi w tych populacjach.”) ale biorąc pod uwagę plany publikacji opracowanych wyników języku Angielskim, będzie zapewne dość czasu i możliwości by zmodyfikować styl pracy, zwłaszcza sekcji wstęp, na bardziej formalny i akademicki.

Przytoczone powyżej drobne uwagi, natury bardziej redakcyjnej niż merytorycznej, nie umniejszają w żadnym stopniu wartości merytorycznej rozprawy i mojej bardzo pozytywnej opinii o tej wyróżniającej się pracy Pani lek. Anny Poradzkiej. W podsumowaniu stwierdzam, że ta bardzo wartościowa rozprawa jest świadectwem, iż Doktorantka wykazała umiejętność samodzielnego rozwiązywania problemów naukowych i badawczych, wykazując potrzebną do tego wiedzę, jak i przygotowanie teoretyczne w zakresie zagadnień, których rozprawa dotyczy. Praca ta zawiera bez wątpienia elementy nowości naukowej, a przedstawione wyniki wskazują, że cele badawcze stawiane przez Doktorantkę zostały zrealizowane z nadwyżką.

Uwzględniając warsztat badawczy, wartości poznawcze i ogromne znaczenie praktyczne dociekań naukowych z dużą przyjemnością przedkładam Wysokiej Radzie Dyscypliny Nauk Medycznych Warszawskiego Uniwersytetu Medycznego wniosek o dopuszczenie lek. Anny Poradzkiej do dalszych etapów przewodu doktorskiego oraz publicznej obrony pracy doktorskiej oraz wyróżnienie pracy.

Kierownik
Zakładu Biostatystyki i Medycyny Translacyjnej
Uniwersytetu Medycznego w Łodzi
Prof. dr hab. n. med. Wojciech Fendler
nr PWZ/2526708
tel. 42 272 53 85 / 691 798 046

Department of Biostatistics
& Translational Medicine

92-215 Lodz, 15 Mazowiecka st.
e-mail: btm @umed.lodz.pl
tel. (042) 272 53 85
www.biostat.umed.pl

