

RECENZJA
rozprawy doktorskiej zatytułowanej:

**„Optimization of Medicine Dosing in Parkinson’s Disease,
Based on Signals from Sensor Measurements”**

autorstwa:

kpt. mgr. inż. Tomasza Gutowskiego

Informacje o rozprawie i jej ogólnej charakterystyka

Rozprawa liczy 183 strony (treść zasadnicza), została podzielona na 8 numerowanych rozdziałów: odpowiednio obszerny wstęp i przegląd stanu wiedzy (rozdziały 1 i 2), omówienie trzech stosowanych zbiorów danych (rozdział 3), ocenę stanu pacjenta (rozdział 4) oraz modelowanie reakcji na dawki leków (rozdział 5) i optymalizację dawkowania (rozdział 6), system informatyczny do śledzenia terapii (rozdział 7) i dyskusję i wnioski (rozdział 8).

Tekst opatrzony jest wykazem bibliograficznym, liczącym 144 pozycje, odpowiednio dobrane do opisywanych treści, poprawnie cytowane w rozprawie. W załączniku zamieszczono wykaz 47 tabel i 63 rysunków. Nie zamieszczono wykazu stosowanych pojęć i skrótów. Spis treści zawiera rozdziały wyłącznie pierwszego i drugiego poziomu, a w pracy pojawiają się jeszcze podrozdziały trzeciego i czwartego poziomu o istotnej objętości i treści, które także powinny być ujęte w spisie treści.

Dyplomant we wstępie nakreśla w sposób opisowy poruszany problem, określane poprzez trzy pytania badawcze, odpowiednio i hierarchicznie prowadzące do rozwiązania problemu:

- wykazanie przydatności zastosowanych technik sensorowych w ocenie stanu pacjenta;
- budowa modeli do predykcji reakcji pacjenta na dawki leku;
- optymalizacja harmonogramu przyjmowania leku.

Takie podejście i zaprezentowane w rozdziałach 4, 5 i 6 wyniki eksperymentów w sposób pełny odpowiadają na postawiony problem badawczy. Brak jednakże formalnego zdefiniowania hipotez badawczych, przykładowo rozdział 2 kończy się stwierdzeniem „the main hypothesis is that [...] the method can create personalized medicine intake schedules [...], co nie jest formalną hipotezą. Dawkowanie może być dobre lub nie, ale wyłącznie w kontekście porównania z czymś, np. dawać istotną poprawę względem tradycyjnego harmonogramu, względem innego modelu opisanego w literaturze.

Rozważyć warto częściową zgodność tytułu rozprawy, w j. polskim określonego jako „Optymalizacja dawkowania leków w chorobie Parkinsona na podstawie sygnałów z pomiarów sensorowych” z jej treścią, skoro tematyka optymalizacji dawkowania poruszana jest tylko w

rozdziale 6 i obejmuje analizę danych syntetycznych pozyskanych z numerycznego modelu farmakokinetyczno-farmakodynamicznego reakcji na dawki leków.

Rozdziały 4, 5 i 6 oparte są częściowo na treściach wcześniej opublikowanych, co bardziej szczegółowo omówione jest w uwagach dotyczących wkładu autorskiego.

Rozdział 7, zawiera szczegóły techniczne implementacji systemu informatycznego do monitorowania terapii. Nie jest on długi i sprawia wrażenie dodanego niepotrzebnie, gdyż trudno w nim doszukać się wątków badawczych.

Kolejne uwagi w recenzji podzielone zostały na uwagi merytoryczne, związane z niedostatecznym zaznaczeniem wkładu autorskiego, uwagi metodologiczne oraz uwagi do układu treści w pracy. Na koniec wymienione są uwagi redakcyjne, które są istotne dla czytelności pracy oraz osobno - drobne uwagi redakcyjne.

Uwagi merytoryczne

Zaznaczenie wkładu autorskiego

W rozprawie przywoływane są cztery prace: [67] Gutowski, Chmielewski 2021, [68] Gutowski 2022, [69] Gutowski Antkiewicz, Szlufik 2023, [77] Gutowski, Chmielewski 2020. Tylko jedna z nich jest indywidualnie napisana przez dyplomanta, natomiast dla pozostałych 3 brakuje określenia jaki był indywidualny wkład autora rozprawy.

Przykładowo rozdział 4 jest istotnym rozwinięciem i częściowym powtórzeniem treści z samodzielnej publikacji autora:

68. Gutowski, T. Deep Learning for Parkinson's Disease Symptom Detection and Severity Evaluation Using Accelerometer Signal. **2022**, 271–276, doi: 10.14428/ESANN/2022.ES2022-107.

Natomiast cały rozdział 5 oparty jest na treści publikacji współautorskiej:

69. Gutowski, T.; Antkiewicz, R.; Szlufik, S. Machine Learning with Optimization to Create Medicine Intake Schedules for Parkinson's Disease Patients. *PLoS One* **2023**, *18*, e0293123, doi:10.1371/JOURNAL.PONE.0293123.

W niej, w sekcji Author Contribution, znaleźć można deklarację, iż T. Gutowski samodzielnie zrealizował Software, Writing – original draft; zaś w pozostałych aspektach współpracowali wszyscy trzej autorzy: Conceptualization, Methodology, Validation, Writing – review & editing.

W rozdziale 6 „Medicine schedules creation” jest przywołana publikacja [65] Watta i in. (2020) w której zaproponowano stosowanie uczenia ze wzmocnieniem do tego samego celu (optymalizacji dawkowania leków). Koniecznie jest wyjaśnienie czy wkład autorski doktoranta polega na dodaniu ograniczeń dla minimalnych odstępów w czasie między dawkami: „To make the model's suggestions more applicable in real-life situations, we have implemented a constraint that imposes a minimum time interval between medication doses”, czy może ten wkład jest szerszy. Wartościowe i rozwiewające te wątpliwości byłoby przytoczenie wprost wyników z pracy [65] jako referencji oraz pokazanie korzyści z modyfikacji zaproponowanej przez doktoranta.

Uwagi metodologiczne

W kilku eksperymentach opisanych w rozprawie nie jest jasna metodologia podziału danych na zbiory treningowy, walidacyjny i testowy. W tabeli 6 scharakteryzowany jest zbiór MUW. Czy grupa 241 badanych została podzielona na podzbiory treningowy, walidacyjny i testowy, w zbalansowany sposób i bez przecieków?

W eksperymencie uczenia głębokiego (str. 46) wykorzystywane są nieprzetworzone dane z trzech osi akcelerometru, jednakże pominięto problem orientacji akcelerometru względem ciała, kończyny pacjenta oraz osi układu odniesienia, np. kierunku przyspieszenia ziemskiego. Używane w ten sposób dane powodować mogą liczne problemy, m.in. nadmierne dopasowanie modelu do nieistotnych cech sygnałów wejściowych. Ponadto sygnały osi X, Y, Z są współzależne a ten sam czujnik umieszczony w inny sposób rejestrować może istotnie odmienne ciągi wartości, nawet w trakcie wykonywania tych samych czynności. Dopiero w rozdziale 4.1.3 Conventional ML Models, autor podejmuje się zadania wyliczania cech z sygnałów i proponuje zastąpienie przebiegów X, Y, Z miarą magnitude (modułu wektora, wzór 20). Wysoce dyskusyjne pozostaje, czy sieci splotowe w wyniku treningu „nabędą” możliwość wydobywania podobnej cechy z danych wejściowych.

Na str. 68 przeczytać można, że z pojedynczego sensora wyliczane są 323 cechy. Powinna być przeprowadzona dyskusja na temat wymiarowości tych danych, w zestawieniu z liczbą rekordów pacjentów w zbiorze treningowym. Czy nie narażamy się w tym wypadku na nadmierne dopasowanie (overfitting)? Należy przemyśleć selekcję cech w każdym eksperymencie, gdzie taki ryzyko istnieje – dyplomant robi to tylko w jednym miejscu.

Diskusyjne jest poleganie na skali oceny od -4 do 4, dla której obie skrajne wartości oznaczają niepożądane objawy, a punkt 0 oznacza stan optymalny. Wartościowy byłby komentarz na temat możliwości rozdzielenia tego problemu na dwa zadania: ocenę dotkliwości dyskinezy i dotkliwości symptomów drżenia. Autor trafnie zauważa niezbalansowanie danych dla obu problemów, wobec czego wskazane wydaje się podjęcie próby niezależnego modelowania dotkliwości symptomów, które jest odpowiednio reprezentowane w zbiorze, zamiast tworzenia jednego modelu dla obu tych zjawisk.

Autor dla wielkości dawki leku oraz dla czasu zastosował standaryzację wartości (str. 104). Dyskusyjne pozostaje czy należy tak zrobić dla wartości mianowanych, których oryginalne skale wartości mają konkretny fizyczny sens i istotny wpływ dla modelowania i dla leczenia.

Dyplomant pisze, że obserwowane są od 10 do 100 krotnie krótsze czasy obliczeń dzięki odpowiedniej implementacji (str. 107). Wydaje się to istotne, ale brak w rozprawie szczegółów, np. dla których modeli to przyspieszenie jest największe.

W trakcie porównania modeli autor zaobserwował, że liczba neuronów w warstwach ukrytych nie ma wpływu na działanie modeli (str. 110). Używano stosunkowo szerokie warstwy: od 16 do 128 neuronów. Czy natomiast zmniejszenie z 16 do 8 lub 4 neuronów będzie miało istotny wpływ na wynik? Czy możliwa jest do określenia minimalna szerokość warstwy nie powodująca istotnego spadku dokładności? Pytania takie zyskują na znaczeniu w momencie rozważania aplikacji docelowej np. działającej w urządzeniu nasobnym o małej pojemności baterii, pamięci, wydajności procesora.

W tabeli 10 (str. 57) zamieszczone są wyniki eksperymentów, jednakże wiele możliwych kombinacji: zadania, zbioru, trybu, mapowania i reprezentacji jest arbitralnie pominiętych. Np.

zbiór SH+SW jest tylko raz w każdym zadaniu i tylko z jedną z możliwych opcji wymiaru wektora zanurzenia (2 lub 4 wymiarowego).

Zabrakło na str. 67 podania wprost wartości N – długości okna analizy Fouriera. Czy wynosi ono 4 sekundy sygnału próbkowanego 50 Hz, czyli 200, wydłużone do najbliższej potęgi liczby 2, czyli do 256?

Autor stosuje 5 metod uczenia maszynowego (str. 68), z czego 3 obsługują parametr `class_weight` dla niezbalansowanych klas. Dla pozostałych powinno stosować się bootstrap albo podpróbkiwanie. Rozważyć można także wzbogacanie (augmentację) danych (jednorazowo wspomniana dopiero na s.119, 5.3.1).

Rozdział 4.1.4 Discussion nie porusza kwestii istotności cech, a mógłby być rozbudowany o krytyczną analizę różnic: czy akcelerometr umieszczony na jednej z kończyn pozwala uzyskiwać lepsze skuteczności, niż na innej, czy wykonywane ćwiczenia wpływają na dokładność, czy wszystkie 323 cechy z jednego czujnika są przydatne w podejmowaniu decyzji przez model?

Autor na str. 80 pisze „both classification and regression models can be used [...] only results for regression are showcased for this dataset, as the balanced metrics for classification are unsatisfactory”, pozostawiając niewyjaśnione wedle jakiego kryterium doszło tu do braku satysfakcjonującej klasyfikacji. Czy było za niskie (jakie) $F1$, precyzja, może czułość?

Na str. 82 znajduje się wyjaśnienie, że każdy eksperyment walidowany jest na dwa sposoby: pozostawianie jednego pacjenta do testów (leave one patient out, LOPO), co jest trafne, gdyż zapobiega przeciekom danych i wykazuje zdolność generalizacji modelu oraz 10 krotną walidację krzyżową (oznaczoną jako 10F), co niestety nie zapobiega przeciekom i niesie ryzyko przypadkowej obecności tych samych pacjentów w zbiorze uczącym i testowym. W ten sposób otrzymać można zawyżone dokładność działania modeli. W rozdziale 4.2.4.1 autor dokumentuje krótko porównanie między dwoma podejściami (tabela 19), ale szczegółowo dalej analizuje tylko wyniki eksperymentu 10F. Komentowane są małe różnice między 10F a LOPO, jednakże wyłącznie na podstawie tabeli 19 jest to nieuzasadnione i nie przeprowadzono testów statystycznych. Może być to wyjaśnione w trakcie obrony.

Model PK/PD zastosowany został do wygenerowania syntetycznych danych dla 50 pacjentów. Cenne byłoby uzasadnienie tej liczby, z czego ona wynikała i jakie byłyby konsekwencje jej podwojenia lub dalszego zwielokrotnienia.

W podsumowaniu Swedish dataset (str. 33) autor pisze „substantial quantity and scope of data has been collected”. Byłoby cenne sprawdzenie i opisanie prac, które wykorzystywały ten zbiór danych, choć zestaw zawiera informacje o 25 osobach i trudno zgodzić się z określeniem „substantial quantity”. Sam autor uczestniczył w badaniach, w których zebrano istotnie większą liczbę danych.

Zrozumiałe jest, że dyplomant korzysta z wiedzy medycznej i standardów obecnych w diagnostyce i leczeniu choroby Parkinsona, jednakże w rozprawie pojawiają się w sposób nieuzasadniony odmienne skale oceny objawów: w zbiorze MUW stosowana jest skala ocen -4...4 lub -10...10, a w innych zbiorach -3...3. Wartościowe byłoby zamieszczenie uzasadnienia dla takiej rozbieżności. Skala ocen Hoehn and Yahr jest tylko wspomniana w pracy (str. 5), ale nie jest wyjaśniona, ani nie zostało uzasadnione, dlaczego autor sam jej nie stosuje.

W rozprawie brakuje użycia kilku pojęć – overfitting, data leakage, patience – które są typowo stosowane w uczeniu maszynowym. Autor na str. 30 słusznie zauważa problemy z dzieleniem

danych ze zbioru MJFF i generalizacją modeli oraz „trained models in the challenge could have learned the fact that certain symptom severities were more likely to occur only with certain patients, during certain sessions and visits”, jednakże rzadko nazywa to przetrenowaniem modelu (overfitting) i nie stosuje terminu przeciek danych (data leakage). Treść się powtarza na str. 42 i 43, gdzie autor pisze o eksperymentach na zbiorze MJFF.

Na str. 44 opisano, że stosowany jest losowy podział, ale udział klas jest podobny w podzbiorach, co nazwać należałoby stratyfikacją. Pojęcie to pojawia się jako: Stratified k-fold cross-validation dopiero na str. 81, gdzie z powodu niewielkiej liczby próbek wykorzystano walidację krzyżową, także stratyfikowaną k-krotną walidację krzyżową. Jest to słuszna metodologia, jednakże wątpliwe jest dlaczego zastosowana została tylko w tym miejscu badań.

Podając na str. 51 szczegóły procesu treningu autor pisze o zatrzymaniu uczenia po 15 epokach w których nie ma obserwowalnej poprawy. Typowo nazywane jest to „cierpliwością”, ale tego terminu nie użyto. Na Fig. 11 i 12 zamieszczono schematy blokowe modeli, ale zabrakło informacji o liczbie uczonych parametrów, o zajętości w pamięci, o czasie treningu pojedynczej epoki oraz o czasie wnioskowania.

Implementację modeli głębokich autor wykonywał samodzielnie: na str. 47 napisane zostało, że w pracy Gutowskiego [68] implementacja wykonana była w bibliotece PyTorch, a późniejsze eksperymenty korzystały z narzędzia TensorFlow. Jeśli autor o tym pisze, to czy ta zmiana jest istotna? Czym ona jest uzasadniona?

Układ treści pracy

W wielu miejscach rozprawy zamieszczane są informacje podstawowe, przegląd stanu wiedzy, wyjaśnienie podstaw wybranych metod. Bardziej właściwe byłoby zebranie tych treści w jednym z pierwszych rozdziałów, a później wyłącznie odwoływanie się do nich. W ten sposób wyraźnie oddzielone byłyby treści opisujące stan wiedzy od wkładu własnego autora rozprawy. W przeglądzie prac innych autorów brak ważnych wyjaśnień, które pozwalałyby na porównanie uzyskiwanych rezultatów: ile klas i jak duże zbiory obejmowały wykonane eksperymenty, raportowane są różne miary, dokładność, AUC, zbalansowana dokładność, nie dające możliwości porównania wyników.

Wyjaśnienie MDS-UPDRS zamieszczone jest na str. 41, ale autor do tej skali oceny odwołuje się wielokrotnie wcześniej (np. na str. 28). Układ treści powinien mieć najpierw podstawy teoretyczne i wyjaśnienia, omówienie metod, wyniki innych autorów, propozycje metod i własne wyniki.

Na str. 52 autor pisząc o wynikach eksperymentów tłumaczy w tym miejscu działanie funkcji aktywacji sigmoid i softmax, co powinno mieć miejsce wcześniej w rozprawie, w tym wzory 5-8, np. na str. 53 wprowadzona jest miara MSE, a na str. 55 w rozdziale Results definiowane są metryki dokładności, czułości, pola pod krzywą ROC. Później jeszcze, na str. 62 zamieszczone są wzory 14-19 z metrykami błędów dla zadań regresji.

Dopiero w rozdziale 4.2.3 (str. 78) autor rozważa selekcję cech i wyjaśnia potrzebę jej stosowania. Objasnia redukcję cech skorelowanych i redukcję wymiarowości LDA, PCA lub selekcję modelem pomocniczym. Jako, że jest to jednym z kluczowych kroków przygotowania danych dla uczenia maszynowego, selekcja cech powinna być omówiona w znacznie wcześniejszych rozdziałach i konsekwentnie stosowana w każdym eksperymencie, w którym zapobiegać trzeba nadmiernemu dopasowaniu modelu z powodu dużej liczby cech i małej liczby próbek.

W rozdziale 4.2.5 na str. 87 autor omawia analizowane dane stosujące skalę TRS, którą dopiero w tym miejscu omawia i wyjaśnia. Coś tak podstawowego dla opisywanej dziedziny powinno być zamieszczone w rozdziałach wprowadzających na początku rozprawy.

Dopiero w rozdziale 5.1. wprowadzany jest model PK/PD lewodopy. Lepsze byłoby umieszczenie tych treści w pierwszej części rozprawy.

Na str. 101-102 wyjaśniane są podstawowe zagadnienia związane z działaniem sieci neuronowych. Lepsze byłoby umieszczenie tych treści w pierwszej części rozprawy.

Uwagi redakcyjne

Poniższe uwagi nie mają wpływu na ocenę pracy ale wpływają na czytelność rozprawy.

Wiele tabel dzielonych jest automatycznie między kolejne strony. W wielu aplikacjach do edycji tekstu istnieją opcje pozwalające np. powtórzyć wiersz nagłówka tabeli na kolejnej stronie lub zapobiegać dzieleniu, co wpłynęłoby pozytywnie na ... Przykładowo na str. 45 umieszczony jest pojedynczy, ostatni wiersz tabeli 7.

Rys. 2 nie ma podanego źródła w podpisie, choć w poprzedzającym akapicie autor wskazuje na [17], brak wartości na osiach, tytuły każdego z trzech wykresów umieszczone są obok osi pionowej, zamiast nad wykresem.

Na str. 24 w rozdziale 2.3 umieszczone są hipotezy i przegląd metod: czy to jest właściwe miejsce na tę treść, a nie wcześniej, gdzie były pytania badawcze?

Rys. 4 nie jest jasny: czy linie przerywane nie oznaczać mają inne relacje niż ciągłe? Kolorowe bloki dostarczają wejścia do białych? Górne białe bloki dostarczają wejścia do kolorowych?

Wyjaśniając działanie sieci splotowej na str. 48 autor pisze o uzupełnianiu zerami (padding), „added around the border of the network” – powinno być „of the image” albo „of the feature maps”.

Komentując wyniki na str. 57 autor pisze o wzroście 1,7%, 6,9% itd. Tymczasem powinno się pisać o punktach procentowych.

W tabeli 14 wzór (24) jest liczony w dziedzinie widma a umieszczony jest w części z nagłówkiem time domain i nie różni się niczym od wzoru (29).

Wyniki w tabeli 15 wprowadziły konwencję wytłuszczenia maksymalnej wartości metryki, która powinna być stosowana konsekwentnie w całej rozprawie.

W tabeli 20 pokazano wyniki dla danych obejmujących więcej niż jedną czynność. W tekście jest mowa o poprawie, ale nie pokazano wprost ile ta poprawa wynosi. Czytelnik może wyliczać różnice między wynikami z tabeli 19 i tabeli 20 lub między różnymi wierszami tabeli 20, ale wykazanie pozytywnych wartości różnic wyników powinno być od razu opracowane w prostej i przystępnej formie.

Wykresy na rys. 25 prezentują histogramy ocen stanu pacjenta przyznawanych przez lekarza oraz przez samego pacjenta. Taki sposób prezentacji danych nie pozwala stwierdzić czy zachodzi zgodność (ten sam pacjent w tym samym momencie ocenia swój stan tak samo jak lekarz), a wyłącznie pokazuje, że liczba ocen jest podobna w całej populacji, co nie jest za

bardzo przydatne. Na osi dolnego wykresu brakuje wartości 1,5, przez co osie obu wykresów są nieproporcjonalne, co jeszcze bardziej utrudniają wyciąganie wartościowych wniosków.

Tytuły tabel 28 oraz 30 są bardzo długie (5 wierszy tekstu) i wyjaśniają użycie testu statystycznego. Taka treść powinna znajdować się w głównym tekście, a tytuł tabeli powinien być zwięzły.

Wzory 53-56 oraz 58, 63, 64 definiują kryteria dla modelu, jednak nie mają lewej strony równania, wobec czego formalnie nie są równaniami.

Drobne uwagi redakcyjne

Str. 5 - niepotrzebnie liczba mnoga „diagnosing is not ... tasks

Str. 5 – literówka „agonists”, chyba „antagonists”

Str. 5 – niejasne zdanie „While brain scans, like dopamine transporter scans...”

Str. 7 – niepotrzebne powtórzenia informacji teoretycznych, np. levodopa is a dopamine precursor...”, które było na str. 5, 6

Str. 8 – nazwa leku levodopa pisana jest niekonsekwentnie z wielkiej litery, gdy tymczasem w pracy przeważa pisownia z małej.

Str. 14 – niepotrzebnie długi tytuł rozdziału 2. The review of the current state of research regarding the application of computer science methods in the management of Parkinson’s disease.

Str. 25 fig 4 oraz str. 64 fig 18 zatyt. „Chart presenting...” – co nie jest właściwym sposobem podpisywania rysunków

Str. 28 fig 5 nie ma podanego źródła.

Str. 29 tabela 1 i 31 tabela 2, 32 tabela 3 nie mają podanego źródła.

Str. 31 tabela 2 zawiera tylko jeden wiersz, co kwestionuje sens podawania informacji w formie tabeli.

Str. 62 literówka „now” – powinno być „not”

Na rys. 26 są umieszczone wykresy punktowe, ze znacznikami wartości na tyle dużymi, że zachodzą na siebie. W tej sytuacji korzystniejsze byłoby użycie wykresu skrzypcowego, tak jak było to na wielu wykresach we wcześniejszych rozdziałach.

Str. 101-102 na rys. 30 i 31 użyte jest oznaczenie TanH zamiast tanh. Podpisy rys. 30 i 31 wydają się nie pasować do ich zawartości.

Str. 104 w tekście pojawia się „this paper” zamiast „this chapter”

Str. 121 jest w tekście odwołanie do table 3 które było na str. 32. Nie byłoby błędem potworzenie tej tabeli w rozdziale 5.3, aby ułatwić czytającemu odbiór treści.

Str. 130 literówka „god metrics”, „good metrics”

Str. 170 w rozdziale 7 przywoływana jest tabela 4, która była na stronie 34. Nie byłoby błędem potworzenie tej tabeli w tym rozdziale, aby ułatwić czytającemu odbiór treści.

Str. 171 ekrany aplikacji prezentowane w tym rozdziale mają raz wersję anglojęzyczną a raz polskojęzyczną.

Źródło 27 to strona internetowa, z podaną datą dostępu marzec 2022. Powinno być to odświeżone, gdyż strona nadal działa.

Wnioski końcowe i podsumowanie recenzji

Rozprawa zawiera wiele wątków badawczych podporządkowanych wspomnianym celom i w ogólności skupia się na podejściach znanych z uczenia maszynowego, w tym na przygotowaniu danych, wyborze modeli, treningu i testowaniu oraz ocenie modeli.

Autor wielokrotnie bardzo trafnie podchodzi do przygotowania danych do eksperymentów, usuwa metadane, które prowadziłyby do nadmiernego dopasowania modelu. Definiuje także trafnie problem oceny symptomów jako zadanie regresji a nie klasyfikacji, co w konsekwencji prowadzi do lepszej możliwości oceny dokładności i pomiaru błędów regresji modeli i stosowanie skali ciągłej. Zaproponował także interesujące podejście do transformacji danych kategoriycznych i numerycznych, które pochodzą z różnych źródeł. Autor rozwiązuje problem połączenia decyzji binarnej: „Tak”, „Nie” oraz skali liczbowej 0...4. Zwraca jednocześnie uwagę na silnie niezbalansowanie zbioru i wskazuje na możliwość stosowania odpowiednich funkcji straty lub próbkowania zbioru. Trafnie opisuje źródło błędów etykiet: subiektywną ocenę przez ekspertów, wynikającą z różnego doświadczenia bądź stronniczości.

Dyplomant stosuje interesujące podejście z zastępowaniem wektorów one-hot reprezentujących wartości zmiennych kategoriycznych poprzez wektory zanurzenia (ang. embedding) w przestrzeni niskowymiarowej o 2, 3, 4 wymiarach i wykazuje korzyść z takiego przekształcenia.

Autor nie poprzestaje na modnym obecnie uczeniu głębokim, ale podejmuje się użycia klasycznych modeli uczenia maszynowego i samodzielnego wyliczenia cech z sygnałów i danych z czujników. Jest to bardzo cennym uzupełnieniem dla stosowania uczenia głębokiego – pozwala na dokładniejsze interpretowanie działania modelu, zgodnie z nurtem XAI (wyjaśnialnej sztucznej inteligencji) i zasługuje na dodatkową pozytywną ocenę.

Autor wyciąga trafne wnioski o możliwości lub trudności w ocenie bradykinezy lub dyskinezy i stopniu drżenia kończyn zależnie od wykonywanych przez pacjentów ćwiczeń. Dostarcza to istotnych praktycznych wniosków.

Podsumowując, praca zawiera samodzielny, istotny i wartościowy wkład udowadniający, że autor dogłębnie poznał problematykę diagnostyki choroby Parkinsona, naturę rejestrowanych sygnałów oraz świadomie stosuje właściwe narzędzia badawcze. Rozprawa stanowi istotny wkład w rozwój dyscypliny Informatyka Techniczna i Telekomunikacja.

Rozprawę doktorską kpt. mgr. inż. Tomasza Gutowskiego oceniam **pozytywnie** i wnioskuję o dopuszczenie do kolejnych etapów przewodu doktorskiego, w tym do jej publicznej obrony rozprawy.

