

# Ocena przydatności statystycznych systemów uczących we wnioskowaniu o stanie sprawności poznawczej pacjentów w podeszłym wieku

## Assessment of the usefulness of statistical learning systems in drawing conclusions about the cognitive performance status of elderly patients

Adam Bednorz<sup>1</sup>, Ewa Lach<sup>2</sup>, Piotr Seiffert<sup>3</sup>

<sup>1</sup> Szpital Geriatryczny im. Jana Pawła II w Katowicach

<sup>2</sup> Wydział Automatyki, Elektroniki i Informatyki, Politechnika Śląska w Gliwicach

<sup>3</sup> Oddział Reumatologii, Szpital Murcki w Katowicach

Neuropsychiatria i Neuropsychologia 2022; 17, 1–2: 83–94

### Adres do korespondencji:

dr n. o zdr. Adam Bednorz

Szpital Geriatryczny im. Jana Pawła II w Katowicach

ul. Morawa 31, 40-353 Katowice

e-mail: adam.bednorz@emc-sa.pl

### Streszczenie

**Wstęp:** Szybki wgląd w stan sprawności poznawczej chorego może w wymierny sposób pomóc w procesie diagnostycznym i zaplanowaniu dalszej opieki nad starszymi pacjentami. Ze względu na stale rosnące wymagania diagnostyczne przydatne może być wykorzystanie statystycznych systemów uczących się związanych z dziedziną uczenia maszynowego. Algorytm uczy się związku między danymi wejściowymi (wynikiem testu) a odpowiadającymi im zmiennymi wyjściowymi (rozpoznanie choroby). Celem pracy była weryfikacja, jak dany model uczenia poradzi sobie z oszacowaniem sprawności poznawczej starszego pacjenta na podstawie danych zgromadzonych w dokumentacji medycznej, a także porównanie wykorzystanych modeli.

**Materiał i metody:** Dokonano retrospektywnego przeglądu badań 280 pacjentów hospitalizowanych w Szpitalu Geriatrycznym im. Jana Pawła II w Katowicach w latach 2015–2019. Dane wejściowe obejmowały wskaźniki biochemiczne, wyniki skal funkcjonalnych (ADL, IADL), choroby współistniejące oraz zmienne socjodemograficzne. Całość danych podzielono w następujący sposób: 2/3 dane do nauki, 1/3 dane testowe. Wykorzystano następujące modele: regresja liniowa, maszyna wektorów nośnych, naiwny klasyfikator Bayesa, metoda k najbliższych sąsiadów.

**Wyniki:** Żaden z modeli nie osiągnął zadowalającej dokładności. Najlepszy wynik uzyskał model regresji liniowej z równomiernym podziałem danych (AUC = 0,57, ACC = 0,60) oraz model regresji liniowej z uczeniem uwzględniającym wszystkie dane (AUC = 0,68, ACC = 0,67).

**Wnioski:** Na podstawie zgromadzonych parametrów nie powinno się budować systemu automatycznego wnioskowania o stanie sprawności poznawczej pacjenta. Model regresji liniowej wymaga kolejnych weryfikacji empirycznych z udziałem liczniejszych i bardziej zróżnicowanych grup.

**Słowa kluczowe:** sprawność poznawcza, otępienie, statystyczne systemy uczące się.

### Abstract

**Introduction:** Quick insight into the patient's cognitive performance can measurably help in diagnostic support and planning further care for elderly patients. Due to the ever-increasing diagnostic requirements, it can be useful to use statistical machine learning systems related to the field of machine learning. The algorithm learns the relationship between input data (test score) and corresponding output variables (diagnosis). The aim of the study was to estimate how a given learning model would cope with the estimation of the cognitive performance of an elderly patient on the basis of data collected in medical records. The aim of the study was also to compare the models used.

**Material and methods:** A retrospective review of the studies of 280 patients hospitalized at the Geriatric Hospital of John Paul II in Katowice in 2015-2019. Input data included biochemical indices, functional scale scores (ADL, IADL), comorbidities and socio-demographic variables. The total data was divided as follows: 2/3 study data, 1/3 test data. The following models were used: linear regression model, support vector machine, naive Bayes classifier, k-nearest neighbors method.

**Results:** None of the models achieved satisfactory accuracy. The best results were obtained from the linear regression model with uniform data division (AUC = 0.57, ACC = 0.60) and the linear regression model with all data learning (AUC = 0.68, ACC = 0.67).

**Conclusions:** Based on the collected parameters, a system of automatic inference about the patient's cognitive performance should not be built. The linear regression model requires further empirical verifications involving larger and more diverse groups on larger and more diverse groups.

**Key words:** cognitive performance, dementia, statistical learning systems

## Wstęp

Pacjenci starsi (wg Światowej Organizacji Zdrowia powyżej 60. roku życia) często mają liczne dolegliwości i schorzenia przewlekłe, co podnosi wskaźnik hospitalizacji i rodzi zapotrzebowanie na opiekę długoterminową. Seniorzy wnoszą do praktyki opieki zdrowotnej złożone historie choroby i scenariusze kliniczne, co sprawia, że konieczne jest opracowanie strategii poprawiających wyniki oceny diagnostycznej i opieki nad pacjentem (Choudhury i wsp. 2020). Jednym z najważniejszych wyzwań w opiece nad pacjentami geriatrycznymi jest postawienie dokładnej i szybkiej diagnozy. Kompleksowa ocena geriatryczna jest jednak procesem czasochłonnym i wymagającym interdyscyplinarnego podejścia. Lekarze podczas wizyt mają ograniczony czas, aby określić występujące problemy z różnych obszarów. Dodatkowe obowiązki wynikające z konieczności prowadzenia dokumentacji klinicznej oraz niedobór lekarzy geriatrów również wpływają na jakość opieki nad pacjentem (Arndt i wsp. 2017). Szybki, jasny i rzetelny wgląd w stan sprawności poznawczej chorego, jako wstępny przegląd, może w wymierny sposób pomóc w procesie diagnostycznym i planowaniu zindywidualizowanej opieki nad pacjentami z wczesnymi zaburzeniami poznawczymi (Bergerona i wsp. 2019).

Otępienie to zespół objawów spowodowanych chorobą mózgu, zwykle o charakterze przewlekłym i postępującym, który dotyka obecnie ponad 40 mln ludzi na całym świecie. Ich liczba wzrosła ponad dwukrotnie od 1990 do 2016 r. (Nichols i wsp. 2019; Prince i wsp. 2013). Identyfikacja pacjentów zagrożonych konwersją deficytów poznawczych do otępienia stanowi jeden z największych priorytetów zdrowia publicznego. Coraz częściej projekty badawcze są ukierunkowane na badania przesiewowe lub predykcję otępienia (Shaik i wsp. 2016; Mendonça i wsp. 2016; Bregman i wsp. 2015). Dysfunkcje poznawcze mogą być oceniane za pomocą różnych testów przesiewowych, na podstawie subiektywnych ocen pacjentów oraz wywiadów z opiekunami (Yim i wsp. 2020). U pacjenta wykazującego wczesne cechy otępienia występuje wiele czynników, które należy wziąć pod uwagę przy interpretacji zmiennego i postępującego deficytu poznawczego. Wyniki badań wskazują na zależności między sprawnością poznawczą a zmiennymi socjodemograficznymi, stanem somatycznym i jakością życia (Barnes i Yafe 2011; Cooper i wsp. 2015; Lyu i wsp. 2014). Inne ważne zmienne w modelach predykcyjnych otępienia to płeć, wskaźnik masy ciała,

ograniczona codzienna aktywność z powodu problemów zdrowotnych, choroby przewlekłe oraz subiektywne samopoczucie (zadowolenie z własnego stanu ekonomicznego oraz zdrowotnego). Na podstawie zgromadzonych danych można założyć istnienie bezpośrednich i pośrednich interakcji między różnymi zmiennymi wpływającymi na funkcje poznawcze (Pinto i wsp. 2016; Rouch i wsp. 2014; Tzang i wsp. 2015). Wspomniane zmienne dostarczają informacji, których nie można uchwycić w prostym teście przesiewowym, takim jak *Mini-Mental State Examination* (MMSE; *Krótką skalą oceny stanu psychicznego*). Zwykła identyfikacja czynników ryzyka lub czynników ochronnych nie jest wystarczająca w przewidywaniu wystąpienia zaburzeń poznawczych i konwersji do otępienia. Wspomniane wyżej zmienne mogą mieć znaczenie w sytuacji, gdy uwzględni się ich wzajemne interakcje z wykorzystaniem odpowiednich algorytmów. W konsekwencji stanowi to wyzwanie dla klinicystów w zakresie oceny funkcji poznawczych i potencjalnych zaburzeń, zwłaszcza w ujęciu długoterminowym. Postęp technologiczny wspierany przez sztuczną inteligencję może ułatwić praktykowi podejmowanie decyzji (Bergerona i wsp. 2019).

Narzędzia sztucznej inteligencji, do których zalicza się uczenie maszynowe (*machine learning* – ML), to technologie pozwalające na analizę danych, wyłapywanie ukrytych informacji, identyfikację zagrożeń dla życia pacjentów czy stawianie diagnoz (Hashimoto i wsp. 2018). W uczeniu maszynowym dostarczane są dane wejściowe (zmienna objaśniająca, wejściowa) razem ze spodziewanymi wynikami (zmienna objaśniana, wyjściowa). Rola oprogramowania polega na poznaniu zasad, które wywołują określone rezultaty poszczególnych danych wejściowych (uczenie nadzorowane) (Koronacki i Ćwik 2005; Raschka 2016).

Algorytm uczy się związku między danymi wejściowymi a odpowiadającymi im zmiennymi wyjściowymi (np. identyfikacja choroby). Po zakończeniu procesu uczenia się algorytm może podawać klasyfikację lub predykcję po wprowadzeniu nowych danych wejściowych (Han i wsp. 2018). Uczenie maszynowe po raz pierwszy zostało zastosowane w medycynie ze względu na jego zdolność do radzenia sobie z dużymi i złożonymi zbiorami danych (Kononenko 2001). Uczenie maszynowe pozwala na ujawnianie znaczących wzorców z dużego, złożonego z wielu zmiennych klinicznych zbioru danych i dalszego „uczenia się” na podstawie praktycznych modeli predykcyjnych. Stwarza to jednocześnie

możliwość dalszego ulepszania modeli poprzez dostarczanie im coraz bardziej zróżnicowanych klinicznie danych o stanie zdrowia pacjenta, np. z zakresu genomiki, biomarkerów, wskaźników behawioralnych oraz dotyczących stylu życia (Alzheimer's Association 2016; Hampel i wsp. 2018). Uczenie maszynowe jest powszechnie stosowane w neuroobrazowaniu w celu wykrycia oznak wczesnej neurodegeneracji w badaniach podłużnych (Gupta i Kahali 2020).

W otępieniu metody ML zastosowano w celu uzyskania wysokiego poziomu dokładności w automatycznej klasyfikacji niesprawności poznawczej danego pacjenta (Salvatore i wsp. 2016). Szeroko wykorzystywano te metody w badaniach neuroobrazowych nad otępieniem oraz do oceny miar poznawczych, behawioralnych i funkcjonalnych (Weakley i wsp. 2015). Algorytmy uczenia maszynowego, takie jak maszyny wektorów nośnych czy drzewa decyzyjne, były stosowane w rozpoznawaniu łagodnych zaburzeń poznawczych (*mild cognitive impairment* – MCI) na podstawie wyników testów przesiewowych (Dyrba i wsp. 2015; Park 2020; Zhang i wsp. 2011). W kilku badaniach użyto ML do diagnostyki różnicowej otępienia. Gurevich i wsp. (2017) wykorzystali dane z Consortium to Establish a Registry for Alzheimer's Disease (CERAD) w celu zidentyfikowania otępienia alzheimerowskiego wśród 158 osób na podstawie biomarkerów płynu mózgowo-rdzeniowego, osiągając dokładność klasyfikacji 89%. Algorytmy uczenia maszynowego, pozwalające na analizę złożonych danych medycznych, wykorzystywano także w diagnostyce neuropsychologicznej (Battista i wsp. 2017; Deo 2015). Weakley i wsp. (2015) zastosowali ML do wybranych testów psychologicznych wykorzystywanych w diagnostyce neuropsychologicznej w grupie 272 pacjentów, aby uzyskać klasyfikację wskaźników klinicznych otępienia. Uwzględniono m.in. *Test łączenia punktów* część A i B, *Test 15 słów Reya*, *Powtarzanie cyfr* i *Symbole cyfr* ze *Skali inteligencji Wechslera dla dorosłych*, *Bostoński test nazywania*, *Test rysowania zegara*, *Geriatryczną skalę oceny depresji*, *TICS (Telephone Interview for Cognitive Status)*, próby oceniające fluencję słowną z baterii Delis-Kaplan, *Brief Visual Memory Test*, test Clox, dodatkowe zmienne socjodemograficzne (wiek, wykształcenie, płeć) oraz skale funkcjonalne (m.in. *Instrumental Activities of Daily Living* – IADL, *Skala oceny złożonych czynności życiowych*). W powyższym badaniu ML pozwoliło na redukcję zmiennych i opracowanie skutecznego modelu predykcyjnego wykorzystującego sześć zmiennych wyodrębnionych z całej baterii testów neuropsychologicznych. W innych

badaniach również zastosowano ML w celu zidentyfikowania testów poznawczych o największej wartości predykcyjnej w rozpoznawaniu otępienia (Gupta i Kahali 2020). Na podstawie ML stwierdzono, które testy psychologiczne były dobrymi predyktorami stanu poznawczego jednostki, przy czym dokładność predykcji na podstawie opracowanego modelu pozostawała na podobnym poziomie dokładności w stosunku do predykcji opartej na całej baterii testów psychologicznych. Podobne podejście zastosowano do oceny funkcji językowych u pacjentów z otępieniem alzheimerowskim (Fraser i wsp. 2015; König i wsp. 2015). Uczenie maszynowe pozwala na uczenie się wzorców z wieloczynnikowych złożonych danych bez opierania się na konwencjonalnych założeniach statystycznych, co sprawia, że odgrywa ono coraz większą rolę w badaniach medycznych (Handelman i wsp. 2018; Ahmed i wsp. 2019).

Reasumując, aby zoptymalizować miary ML, takie jak czułość, swoistość lub dokładność klasyfikacji, w badaniach często starano się rozróżnić zdrowych i „niezdrowych” uczestników lub zidentyfikować pewne choroby przewlekłe. Różne wersje zestawu danych wejściowych (różne funkcje, bazy danych i typy danych) poddawano procedurze uczenia, aby zmaksymalizować wspomniane powyżej miary, a badania zalecały ten model ML, który zapewniał największą dokładność. Celem analiz własnych była weryfikacja, jak dany model poradzi sobie z oszacowaniem sprawności poznawczej w wieku podeszłym na podstawie zgromadzonych danych, a także porównanie wykorzystanych modeli.

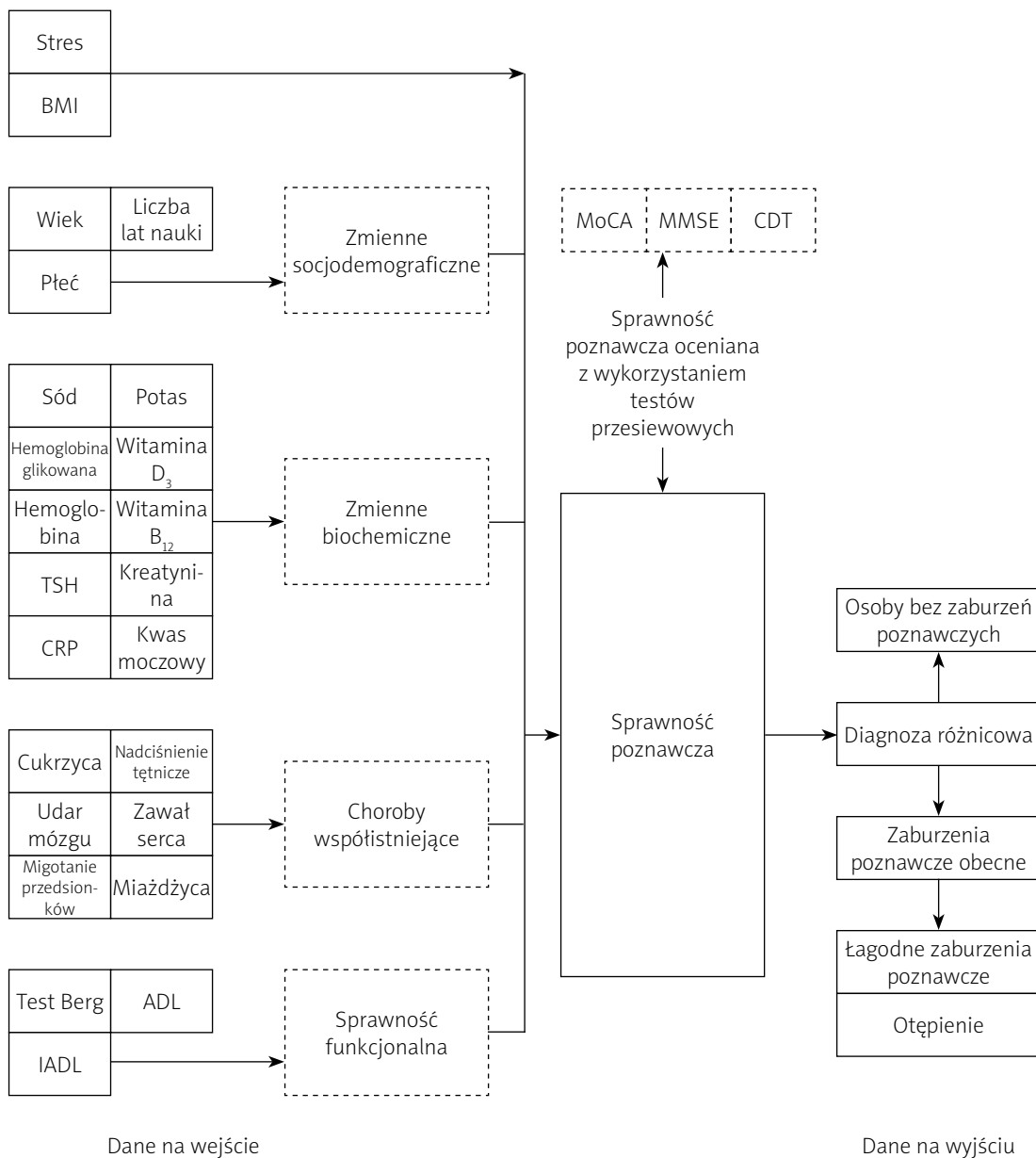
## Materiał i metody

Dokonano retrospektywnego przeglądu badań 280 pacjentów hospitalizowanych w Szpitalu Geriatrycznym im. Jana Pawła II w Katowicach w latach 2015–2019. Wykorzystane dane zostały zgromadzone w ramach przewodów doktorskich. Na przeprowadzenie badań uzyskano zgodę Komisji Bioetycznej ds. Badań Naukowych przy Akademii Wychowania Fizycznego im. Jerzego Kukuczki w Katowicach (Uchwała Nr 2/1/2015 i Uchwała Nr 1/2015). Analizowano przypadki pacjentów z rozpoznaniem łagodnych zaburzeń poznawczych, otępienia bądź pozostających w normie poznawczej. Dane wejściowe obejmowały: zmienne socjodemograficzne (wiek, płeć, liczbę lat nauki), zmienne biochemiczne (stężenie sodu, potasu, kreatyniny, hemoglobiny, białka C-reaktywnego, witamin D i B<sub>12</sub>, kwasu moczowego, hemoglobiny glikowanej,

hormonu tyreotropowego), choroby współistniejące (nadciśnienie tętnicze, cukrzyca typu 2, przebyty zawał serca, migotanie przedsionków, udar mózgu, uogólniona miażdżyca), czynniki stresogenne w przeciągu ostatnich dwóch lat, wskaźnik masy ciała (*body mass index* – BMI), wyniki skal funkcjonalnych *Activities of Daily Living* (ADL; *Skala oceny podstawowych czynności dnia codziennego*) i *Instrumental Activities of Daily Living* (IADL; *Skala oceny złożonych czynności życiowych*), skalę Berg (test zdolności równowagi statycznej i dynamicznej człowieka) oraz skale oceniające sprawność poznawczą pacjenta: MMSE, *Montreal Cognitive Assessment – MoCA*, *Test rysowania zegara* (*Clock Drawing Test* – CDT).

Dane wyjściowe polegały na przyporządkowaniu pacjentów do dwóch grup: 1) zaburzenia poznawcze (MCI i zespół otępienny), 2) osoby bez zaburzeń poznawczych. Schemat modelowania oraz dane na wejściu i wyjściu prezentuje rycina 1.

Do czynników wykluczających udział w badaniu zaliczono: niedosłuch i niedowidzenie uniemożliwiające przeprowadzenie badania zgodnie z protokołem, przebyty uraz czaszkowo-mózgowy, padaczkę, chorobę nowotworową, zdekompensowaną niewydolność krążenia, oddechową i nerek, zaburzenia rozwojowe (ADHD), aktualne leczenie onkologiczne/paliatywne w przebiegu choroby nowotworowej, zaburzenia psychiczne, w tym depresję, schizofrenię paranoidalną, zespół za-



Ryc. 1. Charakterystyka zgromadzonych danych z uwzględnieniem danych wejściowych i wyjściowych

leżności alkoholowej, stosowanie neuroleptyków. Pacjenci, u których ustalono rozpoznanie otępienia i których włączono do dalszych analiz, nie byli wcześniej leczeni prokognitywnie (otępienie po raz pierwszy zdiagnozowano w trakcie hospitalizacji).

W proces gromadzenia danych był zaangażowany interdyscyplinarny zespół specjalistów. Skale funkcjonalne oraz dodatkowe parametry były wypełniane przez pielęgniarkę w izbie przyjęć w dniu przyjęcia pacjenta. W trakcie hospitalizacji psycholog oceniał sprawność poznawczą z wykorzystaniem testów przesiewowych (MMSE, MoCA, CDT), na podstawie których dokonywano późniejszej klasyfikacji (zaburzenia poznawcze są obecne *vs* zaburzenia poznawcze nie występują). Wyniki uzyskane w badaniu psychologicznym nie weszły w skład pierwotnego zestawu danych do nauki i testowania. Skala Berg była wypełniana przez fizjoterapeutę. Dane o współistniejących chorobach somatycznych pochodziły z rozpoznań lekarskich umieszczonych na wypisie szpitalnym.

Ze względu na stosunkowo niewielką liczbę zgromadzonych wyników istotne było wykorzystanie wystarczającej ilości danych w treningu w celu budowy modelu oraz pozostawienie wystarczającej ilości danych do testowania. Całość danych podzielono w następujący sposób: 2/3 dane do nauki, 1/3 dane testowe. Schemat prezentuje rycina 2.

Zestaw danych uczących służy do trenowania różnych modeli, których skuteczność jest następnie sprawdzana. Zaletą wyznaczania zestawu testowego z próbek nieznanymi modelowi na

etapie uczenia oraz doboru modelu jest oszacowanie zdolności uogólniania klasyfikacji.

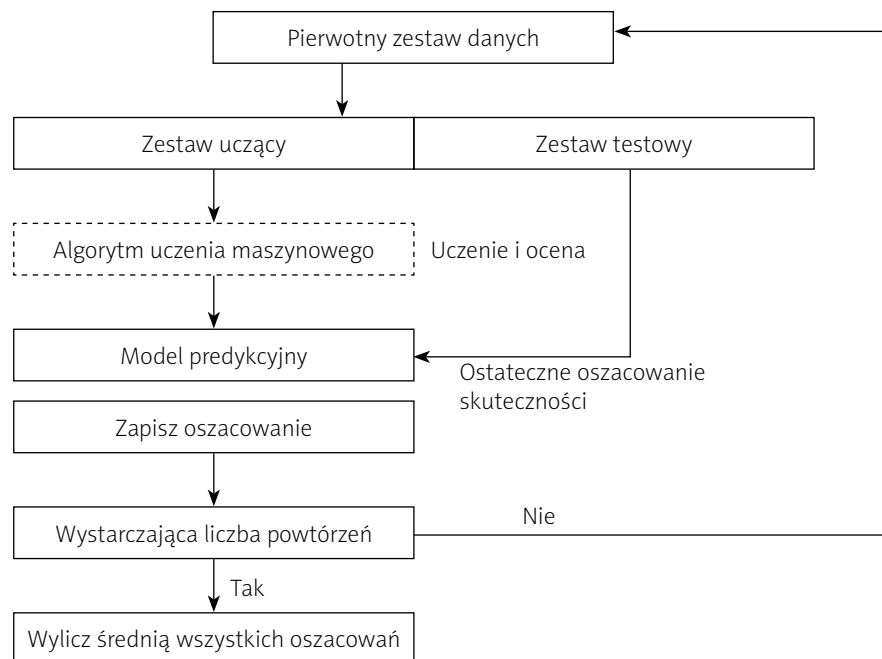
W badanej grupie wykorzystano równomierny podział, który zakładał losowy dobór osób z zaburzeniami oraz bez zaburzeń, tak aby liczebność tych dwóch podgrup była taka sama. Następnie wyniki były dzielone na 2/3 (dane uczące) i 1/3 (dane testowe). Cała procedura była powtarzana 1000 razy.

Początkowo podjęto próbę wydzielenia z każdej grupy (tj. norma/MCI/zespół otępienny) określonej liczby pacjentów, jednak ze względu na niewielką liczbę pacjentów z otępieniem zrezygnowano z tego rozwiązania.

Dodatkowo w kolejnej fazie w celu weryfikacji różnych prób klasyfikacji model był tworzony z zastosowaniem wszystkich danych i następnie na bazie wszystkich zgromadzonych danych testowany.

Do klasyfikacji nie włączono pacjentów z brakującymi danymi. Odstąpiono od uzupełnienia braku danych wartością 0 ze względu na to, że część danych zawierała parametry biochemiczne i z uwagi na możliwe zniekształcenia uzyskanych wyników. Z powodu braku danych dotyczących hemoglobiny glikowanej zrezygnowano z tego parametru. W obliczeniach statystycznych wykorzystano program Matlab (program komputerowy do wykonywania obliczeń naukowych i inżynierskich oraz tworzenia symulacji komputerowych).

Dla wyników wyliczono wartość AUC (*area under the curve*, pole pod krzywą ROC – *receiver*



Ryc. 2. Schemat testowania modeli klasyfikacji

*operating characteristic*, mieszcząca się w przedziale od 0 do 1, im większa, tym lepszy model) oraz ACC (*Accuracy*, tzw. skuteczność reguły decyzyjnej). Krzywa ROC pozwala na ocenę poprawności klasyfikatora i zapewnia łączny opis jego czułości oraz swoistości. AUC to ogólna miara wydajności klasyfikatora dla serii granic. Do określenia skuteczności algorytmu uczenia się wykorzystano macierz pomyłek, w której zliczone zostały wyniki przewidywań klas prawdziwie pozytywnych, fałszywie pozytywnych, prawdziwie negatywnych i fałszywie negatywnych. Poniżej przedstawiono krótką charakterystykę wykorzystanych modeli.

Naiwny klasyfikator Bayesa – przykład klasyfikacji opierającej się na modelu probabilistycznym (prawdopodobieństwo wystąpienia X pod warunkiem wystąpienia Y). Metoda polega na obliczeniu dla każdej obserwacji prawdopodobieństwa jej przynależności do konkretnej klasy. W kolejnym etapie obserwacji przypisuje się do klasy, w której prawdopodobieństwo jest najwyższe (Trevor i wsp. 2009). Przykładowo, pacjent poddaje się testowi na obecność wirusa X. Wiarygodność testu wynosi 90%, a w regionie, w którym mieszka pacjent, nosicielstwo wspomnianego wirusa stwierdzono u 6 na 100 mieszkańców (pytanie: jeżeli wynik pacjenta będzie pozytywny/negatywny, jakie jest prawdopodobieństwo obecności wirusa X w organizmie pacjenta?).

Maszyna wektorów nośnych (*support vector machine* – SVM) – model klasyfikacji, który definiuje się jako odległość między hiperprzestrzenią rozdzielającą (granicą decyzyjną) a najbliższymi próbkami uczącymi, tzw. wektorami nośnymi. SVM maksymalizuje szerokość marginesu pomiędzy dwoma klasyfikacjami (Christianini i Shawe-Taylor 2000; Raschka 2016). W odniesieniu do poruszanej tematyki celem było wskazanie płaszczyzny/hiperpłaszczyzny, która najlepiej rozdzieli punkty na dwa zbiory (zdrowi/chorzy), a następnie sprawdzenie, po jakiej stronie będzie nowy punkt (pacjent).

Metoda k najbliższych sąsiadów (*k nearest neighbours* – kNN) – model klasyfikacji zakładający, że dane należące do tej samej klasy mają podobne wartości cech oraz są położone stosunkowo blisko siebie w przestrzeni. Algorytm kNN uwzględnia interpretację graficzną procesu klasyfikacji. W tym modelu istotne jest zdefiniowanie pojęcia odległości, posługując się odpowiednią metryką (Altman 1992). W odniesieniu do poruszanej tematyki wykorzystano dane zgromadzonej grupy osób z określonymi wynikami (wskaźnikami biochemicznymi, chorobami współistnieją-

mi itd.) oraz poprawną diagnozą, jeśli chodzi o sprawność funkcji poznawczych. Kolejny pacjent był oceniany pod kątem tych wszystkich parametrów, a następnie spośród wcześniejszych przypadków poszukiwano pacjenta najbardziej do niego zbliżonego (założono, że rozpoznanie nowego pacjenta jest takie samo, jak pacjenta najbardziej do niego „zbliżonego”).

Regresja liniowa (*linear regression* – LR) – metoda, której celem jest tworzenie modelu liniowej zależności odpowiedzi (zmiennej Y) w odniesieniu do jednego lub więcej czynników predykcyjnych (zmiennych X objaśniających) (Collett 2002). Do obliczeń wykorzystano uogólniony model regresji liniowej z ustawieniem, że zmienna zależna przyjmuje dwie wartości (chory/zdrowy) i wykorzystano funkcję logit (logitową). Logarytm szansy (lub logit prawdopodobieństwa – stosunek prawdopodobieństwa sukcesu do prawdopodobieństwa porażki) w sposób liniowy zależy od wartości zmiennej objaśniającej.

## Wyniki

Ogólną charakterystykę socjodemograficzną, wyniki badań laboratoryjnych, testów przesiewowych, skal funkcjonalnych oraz pozostałych parametrów przedstawiono w tabeli 1.

Dane dotyczące chorób somatycznych w badanej grupie pacjentów zaprezentowano w tabeli 2.

Po usunięciu wierszy z brakującymi danymi liczba osób z zaburzeniami poznawczymi (MCI lub zespół otępienny) wyniosła 163, a grupy kontrolnej (pacjenci, którzy uzyskali wynik w normie) – 82. W pierwszej kolejności weryfikowano model, opierając się na całości danych z równomiernym podziałem (tj. z grupy osób z zaburzeniami poznawczymi *vs* bez zaburzeń poznawczych wyodrębniono taką samą liczbę pacjentów). Wszystkie modele okazały się nieprzydatne. Model LR osiągnął dokładność 0,6 (AUC = 0,57, ACC = 0,60) w porównaniu z pozostałymi modelami, które były sprawdzane, tj. SVM (AUC = 0,56, ACC = 0,5), kNN (AUC = 0,56, ACC = 0,5), NB (AUC = 0,54, ACC = 0,5). Model regresji liniowej charakteryzował się niskimi wartościami czułości (0,5) i swoistości (0,69). Dane przedstawiono w tabeli 3.

Następnie wszystkie dane wykorzystano do uczenia się, a następnie do testowania. Żaden z weryfikowanych klasyfikatorów nie wykazał zadowalającej dokładności. Model regresji liniowej charakteryzował się skutecznością (AUC = 0,68, ACC = 0,67) w porównaniu z innymi weryfikowanymi modelami – SVM (AUC = 0,5, ACC = 0,34), kNN (AUC = 0,65, ACC = 0,3),

BN (AUC = 0,61, ACC = 0,54). Na podstawie macierzy pomyłek czułość dla modelu regresji liniowej oceniono na 0,82, natomiast swoistość na 0,36. Dane przedstawiono w tabeli 4.

## Dyskusja

Celem pracy była ocena dokładności poszczególnych modeli w klasyfikacji sprawności poznawczej pacjenta na podstawie danych zgromadzonych w dokumentacji medycznej (tj. zmienne socjodemograficzne, wskaźniki biochemiczne,

choroby współistniejące, skale funkcjonalne oraz dodatkowe parametry – BMI, czynniki stresogenne w ostatnich dwóch latach). Celem było skonstruowanie klasyfikatora pozwalającego na oszacowanie, czy pacjent ma zaburzenia poznawcze. Żaden z weryfikowanych klasyfikatorów nie wykazał zadowalającej dokładności. Należy podkreślić, że w części prac dotyczących algorytmów ML testy przesiewowe były uwzględniane w danych do nauki i testowania, w niniejszej pracy zostały one uwzględnione jedynie w przypadku gromadzenia danych oraz podziału na grupy osób

**Tabela 1.** Charakterystyka zmiennych socjodemograficznych, biochemicznych, funkcjonalnych i poznawczych badanych grup pacjentów (M ±SD)

Zmienna	Funkcje poznawcze w normie	Zaburzenia poznawcze	
	Grupa kontrolna (n = 94)	MCI (n = 150)	Otępienie (n = 36)
Wiek	76,3 ±7,8	78,2 ±6,8	81,3 ±3,6
Płeć	77 K (82%)/17 M (18%)	129 K (86%)/21 M (14%)	29 K (81%)/7 M (19%)
Wykształcenie	10,2±2,7	10,2 ±3,6	8,8 ±2,6
MMSE	28,9±1,1	27 ±2,5	21,4 ±3,2
MoCA	27,4 ±1,9	21,7 ±2,8	15,3 ±3,5
CRP	9,7 ±13,4	10,2 ±12,8	11,8 ±18,4
TSH	2,3 ±1,9	2,6 ±2,7	2,5 ±1,9
Witamina B <sub>12</sub>	366,8 ±144,0	397,7 ±223,1	324,8 ±175,7
Natremia (mmol/l)	141,6 ±2,6	141,4 ±3,4	141,4 ±2,7
Kaliemia (mmol/l)	4,4 ±0,4	5,3 ±10,8	4,3 ±0,4
Stężenie kreatyniny w surowicy (mg/dl)	0,9 ±0,2	0,9 ±0,3	0,9 ±0,3
Hemoglobinemia (g/l)	13 ±1,7	13 ±1,4	12,6 ±1,2
Kwas moczowy	5,8 ±1,8	5,6 ±1,4	5,5 ±1,9
Witamina D	21,5 ±13,2	19,2 ±12,3	14,9 ±13,4
Skala ADL	5,2 ±1,1	5,3 ±1,9	4,4 ±1,5
Skala IADL	21,9 ±4,8	21,2 ±4,4	15,8 ±4,8
Skala Berg	39,9 ±11,6	40 ±9,5	35,4 ±9
BMI	29,7 ±5,4	29,2 ±5,3	27,8 ±5,8
Czynniki stresogenne w ostatnich 2 latach	17 – tak, 77 – nie	24 – tak, 126 – nie	9 – tak, 27 – nie

TSH – hormon tyreotropowy, CRP – białko C-reaktywne

**Tabela 2.** Charakterystyka występowania chorób somatycznych w badanych grupach pacjentów (M ±SD)

Choroba	Funkcje poznawcze w normie	Zaburzenia poznawcze		Cała grupa N = 280
	Grupa kontrolna (n = 94)	MCI (n = 150)	Otępienie (n = 36)	
Cukrzyca	36 (38%)	61 (41%)	12 (33%)	109 (39%)
Miażdżyca	24 (26%)	38 (25%)	13 (36%)	126 (45%)
Przebyte zawał serca	5 (5%)	16 (11%)	4 (11%)	25 (9%)
Migotanie przedsionków	12 (13%)	20 (13%)	7 (19%)	39 (14%)
Udar mózgu	6 (6%)	15 (10%)	8 (22%)	29 (10%)
Nadciśnienie tętnicze	72 (77%)	117 (78%)	23 (64%)	212 (76%)

**Tabela 3.** Ocena jakości modeli przy równomiernym podziale danych z uwzględnieniem sprawności poznawczej pacjentów

Model	AUC	ACC	Czułość	Swoistość
LR	0,57	0,60	0,5	0,69
SVM	0,56	0,5	0	1
kNN	0,56	0,5	0	1
NB	0,54	0,5	0	1

LR (linear regression) – regresja liniowa, SVM (support vector machine) – maszyna wektorów nośnych, kNN (k Nearest Neighbours – metoda K najbliższych sąsiadów, NB – Naive Bayes – Naiwny klasyfikator bayesowski

z zaburzeniami poznawczymi i bez zaburzeń poznawczych. W dalszej procedurze nie były one brane pod uwagę, co mogło zmniejszyć skuteczność reguły decyzyjnej (ACC) w porównaniu z wynikami badań zagranicznych (Kang i wsp. 2019; Yim i wsp. 2020). Największą dokładność spośród porównywanych modeli uzyskała regresja liniowa. Model ten może być przydatny w przyszłych badaniach w trakcie weryfikacji klas rozdzielnych liniowo (osoby z zaburzeniami poznawczymi vs osoby bez zaburzeń poznawczych), pozwalając na szacowanie prawdopodobieństwa przynależności do wyznaczonych grup. Jego przydatność potwierdzają wyniki innych badań. W takim ujęciu regresja liniowa może być użyteczna w klasyfikacji wystąpienia danej choroby (tutaj zaburzeń poznawczych) na podstawie znanych parametrów, co często znajdowało już zastosowanie w przypadku danych medycznych (Raschka 2016).

Maszyny wektorów nośnych są bardziej nastawione na punkty znajdujące się najbliższej granicy decyzyjnej (wektory nośne). Należy zaznaczyć, że chociaż SVM nadaje się do klasyfikacji dwóch oddzielnych klas, to w niniejszej pracy jego dokładność pozostawała na niskim poziomie, co mogło wynikać ze specyfiki badanej grupy, której większość stanowili pacjenci w normie i MCI. W przypadku SVM nie udało się wyraźnie rozdzielić jednym kryterium grupy z zaburzeniami poznawczymi i bez zaburzeń poznawczych na podstawie zgromadzonego zestawu danych i parametrów.

Model najbliższych sąsiadów (kNN) również cechował się niską dokładnością. Możliwe, że osoby z grupy z zaburzeniami poznawczymi i grupy kontrolnej były do siebie zbyt podobne, jeśli chodzi o uwzględnione parametry. W przypadku kNN parametry, które zostały zgromadzone w niniejszej pracy, mogły nie być reprezentatywne dla rozdzielenia grup (osoby z zaburzeniami poznawczymi vs osoby bez zaburzeń poznawczych). Tutaj należy również

**Tabela 4.** Ocena jakości weryfikowanych modeli z uwzględnieniem wszystkich danych

Model	AUC	ACC	Czułość	Swoistość
LR	0,68	0,67	0,82	0,37
SVM	0,5	0,34	0	1
kNN	0,66	0,34	0	1
NB	0,61	0,54	0,37	0,85

LR – Logistic Regression – regresja liniowa, SVM – Support Vector Machine – maszyna wektorów nośnych, kNN (k nearest neighbours) – metoda najbliższych sąsiadów, NB (naive Bayes) – naiwny klasyfikator bayesowski

wskazać na inne ograniczenie algorytmu kNN, który jest wrażliwy na przetrenowanie z powodu wielowymiarowości (przestrzeń cech wraz ze wzrostem liczby wymiarów zestawu danych uczących się o ustalonym rozmiarze staje się coraz bardziej rozległa, tj. możliwa jest sytuacja, że nawet najbliżsi sąsiedzi mogą się znajdować zbyt daleko, aby uzyskać za ich pomocą dobre oszacowanie).

Naiwny klasyfikator bayesowski stanowi prostą metodę uczenia się, którą trudno zakłócić nieistotnymi szczegółami. Metoda ta jest użyteczna w sytuacji, gdy zasilana jest duża ilością danych. W pracy ich ilość mogła być niewystarczająca, stąd niska dokładność tego modelu.

Reasumując – uzyskana niska skuteczność weryfikowanych modeli nie może dyskredytować tego typu rozwiązań i prób ich dalszego zastosowania jako uzupełnienia i wsparcia w procesie diagnostycznym, zwłaszcza że badania zagraniczne potwierdzają ich użyteczność (Na 2019). Należy podkreślić, że statystyczne systemy uczące w żaden sposób nie powinny zastąpić przesiewowej oceny funkcji poznawczych, są one raczej dodatkowym etapem procesu diagnostycznego ułatwiającym wcześniejszą identyfikację pacjentów w podeszłym wieku, u których warto wykonać przesiewową ocenę funkcji poznawczych. Takie rozwiązanie może być przydatne zwłaszcza w sytuacji nadmiernego obciążenia pracą i ograniczeń czasowych, jakie występują u lekarzy pierwszego kontaktu.

W pracy większość testowanych modeli okazała się nieprzydatna, co może wynikać z ich specyfiki, wspomnianej wcześniej. Istotna jest również niewielka liczba osób badanych, w badaniach zagranicznych modele były weryfikowane na większych grupach pacjentów (Lyu i wsp. 2014). Należy jednak zauważyć, że dostarczenie większej liczby tych samych danych może nie zwiększyć dokładności modelu ze względu na zbyt dużą złożoność weryfikowanego obszaru, tj. sprawności poznawczej. Różnicowanie po-



między osobami z zaburzeniami poznawczymi i bez zaburzeń nie jest kwestią algorytmu, ale zależy od wnioskowania klinicznego, które musi uwzględniać kryteria diagnostyczne oraz patomechanizm. Warto podkreślić, że metody uczenia maszynowego mają na celu przede wszystkim klasyfikację i nie budują automatycznie hipotez dotyczących związku danych wejściowych z danymi wyjściowymi (Glaser i wsp. 2019). Niniejsza praca, mimo niskiej dokładności testowanych modeli, może stanowić sugestię odnośnie do kierunku przyszłych badań oraz uwzględnianych w nich parametrów i algorytmów uczenia maszynowego.

Wśród ograniczeń pracy należy wskazać na charakter zebranych danych, które pochodziły z przewodów doktorskich i z tego względu były obciążone licznymi kryteriami wykluczającymi, wskazanymi we wcześniejszej części artykułu. Niewątpliwie w przyszłych badaniach należy uwzględnić zdecydowanie większą liczbę chorób współistniejących oraz podać szacunkowy czas ich trwania. W niniejszej pracy stanowiło to jedno ze wstępnych założeń, nie zawsze jednak była możliwość uzyskania takich informacji z wywiadu i dokumentacji medycznej.

Istotna pozostaje także kwestia podziału danych na podzbiory uczące i testowe, ponieważ w zależności od zastosowanej proporcji następuje rezygnacja z części cennych informacji, z których algorytm uczenia się mógłby skorzystać. W praktyce proporcje pomiędzy danymi uczącymi a testowymi najczęściej wynoszą 60 : 40, 70 : 30, 80 : 20, w zależności od początkowego rozmiaru zestawu. W przypadku dużych zbiorów danych często stosuje się również proporcje 90 : 10 czy 99 : 1 (Raschka 2016). W przyszłych badaniach do rozważenia jest przetestowanie innych podziałów danych na podzbiory uczące i testowe.

Wśród innych ograniczeń należy podkreślić, że klasyfikacja binarna (choroba występuje lub nie występuje) niekoniecznie pozwala w pełni uwzględnić złożoność problemów zdrowotnych u pacjentów geriatrycznych (trafniejsze byłoby pewnego rodzaju kontinuum zdrowia i choroby). Dane wejściowe powinny też uwzględniać farmakoterapię. Ocena polipragmazji (wielolekowość, przyjmowanie co najmniej 5 leków jednocześnie) wydaje się szczególnie istotna u osób starszych (Rochon i wsp. 2021). Szacuje się, że ponad 1/3 osób pomiędzy 75. a 85. rokiem życia stosuje przynajmniej 5 leków (Dagli i Sharma 2014). Wśród nich znajduje się wiele mogących negatywnie oddziaływać na funkcje poznawcze. Do grup leków mających najbardziej istotny wpływ zalicza się m.in. leki o właściwościach

antycholinergicznym, które poprzez działanie na receptory acetylocholiny w ośrodkowym i obwodowym układzie nerwowym zmniejszają jej ilość w synapsach nerwowych. Do oceny ich wpływu można wykorzystać specjalnie stworzone skale (Sørensen i wsp. 2021). Należy uwzględnić także efekt leków stosowanych w terapii otępienia, takich jak inhibitory acetylocholinoesterazy czy memantyna. Wpływ na funkcje poznawcze leków z innych grup, np. przeciwpsychotycznych, leków na nadciśnienie tętnicze czy przeciwkrzepialnych, jest tematem wielu publikacji i aktualnych badań. Wskazane byłoby także uwzględnienie interakcji pomiędzy lekami, które również mogą wpływać na funkcje poznawcze starszych pacjentów. W przyszłych badaniach przydatne będą także dane z badań neuroobrazowych, ocena funkcji poznawczych dokonywana przez bliskich pacjenta, wywiad rodzinny (obciążenia genetyczne w kierunku otępienia, długowieczność w rodzinie), ocena w kierunku podejrzenia zespołu kruchości (zmęczenie, utrata masy ciała), wskaźniki dotyczące wcześniejszych okresów życia (wydarzenia traumatyczne, przewlekłe czynniki stresogenne) oraz aktualnej sytuacji społecznej (samotność, wsparcie społeczne). Na podstawie przeglądu piśmiennictwa stwierdzono, że pożądana może być identyfikacja również mniej oczywistych czynników, takich jak: narażenie na szkodliwe czynniki biologiczne, chemiczne i fizyczne w warunkach domowych i w pracy (np. grzyby/pleśń, praca w przemyśle metalurgicznym), używki (nikotynizm, nadużywanie alkoholu), uraz czaszkowo-mózgowy w wywiadzie czy niektóre rzadziej diagnozowane choroby (np. obturacyjny bezdech senny, przewlekłe zapalenie zatok, choroby przyzębia, zaparcia) (Bredesen 2018). Duża liczba zmiennych wymusza organizację bardziej kompletnego zbierania danych już na poziomie izby przyjęć. Algorytmy uczenia ML stwarzają szansę na wyodrębnienie zmiennych, które odznaczają się największą wartością predykcyjną w przypadku otępienia. W kolejnych badaniach testowane modele powinny być również weryfikowane w kierunku bardziej dokładnej klasyfikacji sprawności poznawczej, jak prezentują to badania zagraniczne, tj. norma/MCI/zespół otępienny (w niniejszej pracy podział: zaburzenia poznawcze obecne vs osoby bez zaburzeń poznawczych) (Yim i wsp. 2020).

Zakres przeprowadzonego badania pozostaje ograniczony, ponieważ oceniano dane z jednego ośrodka badawczego. Możliwość uogólniania modelu hipotetycznego zależy od inkluzywności zbioru danych uczących. Oznacza to, że

hipotetyczny model, który został wytrenowany przy użyciu danych z jednego ośrodka lub małej próby, może nie nadawać się do zastosowania w przypadku osób z innej populacji lub dużej próby. Nauka modelu i tworzona na tej podstawie klasyfikacja odbywa się w oparciu o zbiór danych. Podobnie jak w przypadku mózgu, który nie posiada niezmiennej wiedzy, ale stale przyswaja nowe informacje, również sztuczna inteligencja podlega podobnym procesom przyswajania nowych danych. Im więcej danych zostanie zapewnionych na wejściu, tym model uzyskuje większą liczbę wzorów kombinacji między zmiennymi, co przekłada się na dokładność jego klasyfikacji. Z tego względu na poziomie organizacyjnym istotne jest tworzenie sieci współpracujących ze sobą ośrodków (szpitale, poradni zaburzeń pamięci), gromadzących i wymieniających dane z uwzględnieniem pomiarów na przestrzeni lat, co pozwala na monitorowanie dynamiki zmian sprawności poznawczej.

Obecne aplikacje wykorzystujące ML są ograniczone zarówno przez dostępność danych, jak i zdolność algorytmów do wysuwania z nich wniosków. Większość zastosowań uczenia maszynowego do analiz elektronicznych danych zdrowotnych pacjentów wykorzystuje techniki uczenia pod nadzorem do przewidywania określonych punktów końcowych (Collins i wsp. 2015; Rajkomar i wsp. 2018; Miotto i wsp. 2016; Choi i wsp. 2016). Sama analiza brakujących danych bądź wykorzystanie ilości lub charakteru brakujących danych jako zmiennej wejściowej również może stanowić interesujący obszar badań (Weakley i wsp. 2015). W przyszłych pracach przydatne będzie uwzględnienie w analizach ML, oprócz wyników ogólnych testów przesiewowych (np. MMSE, MoCA), także poziomu wykonania poszczególnych zadań wchodzących w ich skład (np. wynik odroczonego przypominania, zapamiętywania, funkcji wzrokowo-przestrzennych, funkcji językowych). W niektórych badaniach na tej podstawie tworzono całą domenę poznawczą (tj. na ocenę pamięci składały się wyniki poszczególnych prób oceniających funkcje mnesticzne z MMSE, MoCA) lub innych wybranych testów neuropsychologicznych, takich jak *Kalifornijski test uczenia się językowego (California Verbal Learning Test – CVLT)* (Battista i wsp. 2017; Weakley i wsp. 2015).

Badania naukowe wskazują, że uczenie maszynowe może być przydatne do przeprowadzania statystycznie znaczących obliczeń i wspomagania klinicystów w wyborze najlepszych środków (i pomiarów), zmniejszając jednocześnie subiektywność i czasochłonność procesu diagnostycz-

nego. Z tego względu dziedzina ta może być przydatna w optymalizacji obecnych testów psychologicznych lub w projektowaniu nowych testów do oceny poznawczej przyszłych generacji seniorów (Weakley i wsp. 2015). Pandemia, która często ograniczała możliwość rozpoczynania nowych projektów badawczych, paradoksalnie może się okazać szansą na pogłębioną analizę danych już zgromadzonych w dokumentacji medycznej w szpitalnych archiwach. W takim ujęciu nowe spojrzenie na stare dane może się okazać rozwojowe w dalszej perspektywie. Długofalowym celem może być zaimplementowanie określonego modelu do aplikacji stacjonarnej lub mobilnej w postaci platformy diagnostycznej, która po otrzymaniu danych wejściowych (np. po wpisaniu wyników badań laboratoryjnych, skali funkcjonalnych lub w wyniku automatycznej integracji różnych zbiorów danych) z pewnym prawdopodobieństwem szacowałaby sprawność poznawczą danego pacjenta oraz identyfikowała pacjenta wymagającego poszerzonej diagnostyki przesiewowej lub neuropsychologicznej w celu potwierdzenia wcześniejszej predykcji.

## Wnioski

Na podstawie zgromadzonych w badaniu parametrów nie powinno się budować systemu automatycznego wnioskowania o stanie sprawności poznawczej pacjenta. Model regresji liniowej może się okazać użyteczny w przyszłych badaniach, ale wymaga kolejnych weryfikacji empirycznych z udziałem liczniejszych i bardziej zróżnicowanych grup.

## Oświadczenie

Autorzy zgłaszają brak konfliktu interesów.

## Piśmiennictwo

- Ahmed MR, Zhang Y, Feng Z i wsp. Neuroimaging and machine learning for dementia diagnosis: recent advancements and future prospects. *Rev Biomed Eng* 2019; 12: 19-33.
- Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* 1992; 46: 175-185.
- Alzheimer's Association. 2016 Alzheimer's disease facts and figures. *Alzheimers Dement* 2016; 12: 459-509.
- Arndt BG, Beasley JW, Watkinson MD i wsp. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. *Ann Fam Med* 2017; 15: 419-426.
- Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *Lancet Neurol* 2011; 10: 819-828.

6. Battista P, Salvatore C, Castiglioni I. Optimizing neuropsychological assessments for cognitive, behavioral, and functional impairment classification: a machine learning study. *Behav Neurol* 2017; 2017: 1850909.
7. Bergerona MF, Landsetb S, Tarpin-Bernardc F i wsp. Episodic-memory performance in machine learning modeling for predicting cognitive health status classification. *J Alz Dis* 2019; 70: 277-286.
8. Bredesen DB. Wygraj z alzheimerem. Wydawnictwo Galaktyka, Łódź 2018.
9. Bregman N, Regev K, Moore O i wsp. A simple tool to reach populations at risk for developing dementia and Alzheimer's disease. *J Alzheimers Dis* 2015; 46: 151-155.
10. Choi E, Bahadori MT, Schuetz A i wsp. Doctor AI: predicting clinical events via recurrent neural networks. *JMLR Workshop Conf Proc* 2016; 56: 301-318.
11. Choudhury A, Renjilian E, Asan O. Use of machine learning in geriatric clinical care for chronic diseases: a systematic literature review. *JAMIA Open* 2020; 3: 459-471.
12. Christianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge UK 2000.
13. Collett D. Modeling binary data. Chapman and Hall, New York 2002.
14. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015; 372: 793-795.
15. Cooper C, Sommerlad A, Lyketsos CG i wsp. Modifiable predictors of dementia in mild cognitive impairment: a systematic review and meta-analysis. *Am J Psychiatry* 2015; 172: 323-334.
16. Dagli RJ, Sharma A. Polypharmacy: a global risk factor for elderly people. *J Int Oral Health* 2014; 6: i-ii.
17. Deo RC. Machine learning in medicine. *Circulation* 2015; 132: 1920-1930.
18. Dyrba M, Barkhof F, Fellgiebel A i wsp. Predicting prodromal Alzheimer's disease in subjects with mild cognitive impairment using machine learning classification of multimodal multicenter diffusion-tensor and magnetic resonance imaging data. *J Neuroimaging* 2015; 25: 738-747.
19. Fraser KC, Meltzer JA, Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech. *J Alzheimers Dis* 2016; 49: 407-422.
20. Glaser JI, Benjamin AS, Farhoodi R i wsp. The roles of supervised machine learning in systems neuroscience. *Prog Neurobiol* 2019; 175: 126-137.
21. Gupta A, Kahali B. Machine learning-based cognitive impairment classification with optimal combination of neuropsychological tests. *Alzheimer's Dement* 2020; 6: e12049.
22. Gurevich P, Stuke H, Kastrop A i wsp. Neuropsychological testing and machine learning distinguish Alzheimer's disease from other causes for cognitive impairment. *Front Aging Neurosci* 2017; 9: 114.
23. Hampel H, O'Bryant SE, Molinuevo JL i wsp. Blood-based biomarkers for Alzheimer disease: mapping the road to the clinic. *Nat Rev Neurol* 2018; 14: 639-652.
24. Han SH, Kim KW, Kim S i wsp. Artificial neural network: understanding the basic concepts without mathematics. *Dement Neurocogn Disord* 2018; 17: 83-89.
25. Handelman GS, Kok HK, Chandra RV i wsp. eDoctor: machine learning and the future of medicine. *J Intern Med* 2018; 284: 603-619.
26. Hashimoto DA, Rosman G, Rus D i wsp. Artificial intelligence in surgery: promises and perils. *Ann Surg* 2018; 268: 70-76.
27. Kang MJ, Kim SY, Na DL i wsp. Prediction of cognitive impairment via deep learning trained with multi-center neuropsychological test data. *BMC Med Inform Decis Mak* 2019; 19: 231.
28. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med* 2001; 23: 89-109.
29. Koronacki J, Ćwik J. Statystyczne systemu uczące się. Wydawnictwo Naukowo-Techniczne, Warszawa 2005.
30. König A, Satt A, Sorin A i wsp. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimers Dement (Amst)* 2015; 1: 112-124.
31. Lyu J, Lee CM, Dugan E. Risk factors related to cognitive functioning: a cross-national comparison of U.S. and Korean older adults. *Int J Aging Hum Dev* 2014; 79: 81-101.
32. Mendonça MD, Alves L, Bugalho P. From subjective cognitive complaints to dementia: who is at risk? a systematic review. *Am J Alzheimers Dis* 2016; 31: 105-114.
33. Miotto R, Li L, Kidd BA i wsp. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep* 2016; 6: 26094.
34. Na KS. Prediction of future cognitive impairment among the community elderly: A machine-learning based approach. *Sci Rep* 2019; 9: 3335.
35. Nichols E, Szoeke CE, Vollset SE i wsp. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990-2016: a systematic analysis for the global burden of disease study 2016. *Lancet Neurol* 2019; 18: 88-106.
36. Park JH. Machine-learning algorithms based on screening tests for mild cognitive impairment. *Am J Alzheimers Dis Other Dement* 2020; 35: 1533317520927163.
37. Pinto JM, Fontaine AM, Neri AL. The influence of physical and mental health on life satisfaction is mediated by self-rated health: A study with Brazilian elderly. *Arch Gerontol Geriatr* 2016; 65: 104-110.
38. Prince M, Bryce R, Albanese E i wsp. The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimers Dement* 2013; 9: 63-75.
39. Rajkomar A, Oren E, Chen K i wsp. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* 2018; 1: 18.
40. Raschka S. Python uczenie maszynowe. Wydawnictwo Helion, Gliwice 2016.
41. Rochon PA, Petrovic M, Cherubini A i wsp. Polypharmacy, inappropriate prescribing, and deprescribing in old age: through a sex and gender lens. *Lancet Healthy Longev* 2021; 2: e290-300.
42. Rouch I, Achour-Crawford E, Roche F i wsp. Seven-year predictors of self-rated health and life satisfaction in the elderly: the PROOF study. *J Nutr Health Aging* 2014; 18: 840-847.
43. Salvatore A, Cerasa P, Battista MC i wsp. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. *Front Neurosci* 2015; 9: 307.
44. Shaik MA, Khoo CH, Thiagarajah AG i wsp. Pilot evaluation of a dementia case finding clinical service using the informant AD8 for at-risk older adults in primary health care: a brief report. *J Am Med Dir Assoc* 2016; 17: 673.e5-673.e8.
45. Sørensen SR, Frederiksen JD, Anru PL i wsp. Use of drugs with anticholinergic properties at hospital admission associated with mortality in older patients: a Danish nationwide register-based cohort study. *Drugs Real World Outcomes* 2022; 9: 129-140.

46. Trevor H, Tibshirani R, Friedman F. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. Springer Series in Statistics 2009.
47. Tzang RF, Yang AC, Yeh HL i wsp. Association of depression and loneliness with specific cognitive performance in non-demented elderly males. *Med Sci Monit* 2015; 21: 100-104.
48. Weakley A, Williams JA, Schmitter-Edgecombe M i wsp. Neuropsychological test selection for cognitive impairment classification: a machine learning approach. *J Clin Exp Neuropsychol* 2015; 37: 899-916.
49. Yim D, Yeo TY, Park MH. Mild cognitive impairment, dementia, and cognitive dysfunction screening using machine learning. *J Int Med Res* 2020; 48: 1-10.
50. Zhang D, Wang Y, Zhou L i wsp. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 2011; 55: 856-867.