

Ewelina Słowik

e-mail: 189398@ue.wroc.pl

ORCID: 0009-0007-9680-3623

Uniwersytet Ekonomiczny we Wrocławiu

## Zastosowanie przetwarzania języka naturalnego w procesie diagnozowania i leczenia choroby Alzheimerera

DOI: 10.15611/2024.80.2.07

JEL Classification: C61

@ 2024 Ewelina Słowik

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>

**Cytuj jako:** Słowik, E. (2024). Zastosowanie przetwarzania języka naturalnego w procesie diagnozowania i leczenia choroby Alzheimerera. W: H. Dudycz (red.), *Informatyka w biznesie* (s. 92-104). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

**Streszczenie:** W niniejszym artykule podjęto się analizy roli, jaką odgrywają różne techniki przetwarzania języka naturalnego (*Natural Language Processing* – NLP) w diagnostyce oraz leczeniu choroby Alzheimerera. Ponadto przyjrano się interakcji pomiędzy lingwistyką a neurobiologią w kontekście wykorzystania tych technik w medycynie. Metody badawcze zastosowane w artykule obejmują przegląd literatury, analizę porównawczą określonych badań oraz identyfikację obszarów do dalszych prac i propozycje ulepszeń zastosowania technik przetwarzania języka naturalnego w schorzeniu Alzheimerera. Na podstawie analizy porównawczej badań stwierdzono, iż istnieje konieczność dalszego opisu zagadnienia, rozwoju NLP z wykorzystaniem uczenia głębokiego, personalizacji leczenia, stworzenia globalnych baz danych oraz interdyscyplinarnej współpracy dla poprawy korzyści dla pacjentów, personelu medycznego i systemów opieki zdrowotnej.

**Słowa kluczowe:** przetwarzanie języka naturalnego, NLP, choroba Alzheimerera

### 1. Wstęp

Od wielu lat obserwuje się na świecie zjawisko starzenia się społeczeństw oraz wzrost odsetka osób w podeszłym wieku. Prowadzi to do sytuacji, w której coraz więcej z nich cierpi na różnego rodzaju choroby neurologiczne. Statystyki wskazują, iż w 2005 roku w Polsce było około 6 milionów osób powyżej 60. roku życia, a w 2030 roku będzie już około 9 milionów, których te schorzenia dotyczą (Gaweł i Potulska-Chromik, 2015). Choroby neurologiczne to schorzenia zaburzające prawidłowe funkcjonowanie układu nerwowego, który składa się z ośrodkowego układu nerwowego (OUN) oraz obwodowego układu nerwowego odpowiedzialnego za sterowa-

nie aktywnością całego organizmu (Poradnik Gemini, b.d.). Wśród tych chorób można wyróżnić m.in. schorzenia neurodegeneracyjne, które charakteryzują się zaburzeniami funkcjonowania lub umieraniem komórek należących do ośrodkowego układu nerwowego (mózgowie i rdzeń kręgowy) (National Cancer Institute, b.d.). Ośrodkowy układ nerwowy odpowiada np. za regulację pracy narządów wewnętrznych, proces myślenia, zapamiętywania, analizę bodźców docierających z otoczenia, kontrolę napięcia mięśni (Thau i in., 2022). Wraz ze starzeniem się choroby te postępują i często nie ma na nie lekarstwa. Do nich można zaliczyć m.in.: chorobę Alzheimera, chorobę Parkinsona, rdzeniowy zanik mięśni oraz stwardnienie rozsiane.

W przypadku tych schorzeń poprawna diagnostyka jest niezbędna, gdyż umożliwia wiarygodniejsze dalsze prognozowanie choroby oraz przyczynia się do włączenia specyficznego leczenia (Erkinen i in., 2018). W związku z tym, iż wczesne wykrycie za pomocą badań przesiewowych umożliwia wdrożenie leczenia przed wystąpieniem ciężkich objawów klinicznych (Shusharina i in., 2023), w ostatnich kilkudziesięciu lat ocena kliniczna schorzeń neurodegeneracyjnych stała się głównym przedmiotem badań (Dubois i in., 2016; Wu i in., 2011). Do niej można zaliczyć m.in. biomarkery krwi i płynu mózgowo-rdzeniowego, dane z neuroobrazowania, jak PET czy MRI, oraz wyniki badań neurologicznych. Zatem liczba technik, które mogą być wykorzystywane do rozpoznawania schorzeń neurodegeneracyjnych, jest duża. Jednakże należy nadmienić, iż większość tych chorób charakteryzuje się niską swoistością odnośnie danych markerów oraz technik diagnozowania (Shusharina i in., 2023). Oznacza to, iż zarówno metody, jak i markery nie są wystarczająco precyzyjne – zbyt ogóle i nie pozwalają na dokładne rozróżnienie między różnymi chorobami. Często prowadzi to do konieczności postawienia diagnozy przy użyciu wielu metod oraz kosztownych badań (Tagaris i in., 2018). Leczenie schorzeń neurodegeneracyjnych ze względu na nieodwracalny charakter zwyrodnienia tkanki nerwowej polega jedynie na ich spowolnieniu oraz poprawie jakości życia osób z tymi zaburzeniami (Shusharina i in., 2023).

Warto zauważyć, iż zaburzenia zdolności komunikacji w chorobach neurodegeneracyjnych są częste, dlatego zaczęto interesować się wydajnością językową, która obejmuje wypowiedanie słów oraz zrozumienie słów. W ostatnich latach nagromadzono bardzo dużo danych medycznych od profilów genetycznych aż po dane z obrazowania jak MRI (Myszczynska i in., 2020). Dodatkowo rozwinęła się znacznie dziedzina sztucznej inteligencji w tym metody uczenia maszynowego, głębokiego oraz przetwarzanie języka naturalnego. Jednocześnie rozwój sztucznej inteligencji, w tym techniki uczenia maszynowego, głębokiego uczenia oraz przetwarzania języka naturalnego, otworzył nowe możliwości analizy tych danych. W odpowiedzi na to sztuczna inteligencja zaczęła być wykorzystywana do analizy różnorodnych danych medycznych, zwłaszcza że wiele z nich jest obarczonych szumem, który utrudnia wydobycie istotnych informacji biologicznych. Rosnąca liczba przypadków choroby Alzheimera i innych chorób neurodegeneracyjnych wymaga poszukiwania skutecznych metod diagnozowania i leczenia.

Celem niniejszego artykułu jest zbadanie potencjału zastosowania technik przetwarzania języka naturalnego (NLP) w procesie diagnozowania i leczenia choroby Alzheimera, co pomoże zrozumieć, jakie istnieją interakcje pomiędzy lingwistyką a neurobiologią w kontekście tego schorzenia.

Wykorzystano takie metody badawcze jak: przegląd literatury, analizę porównawczą oraz wkład własny – identyfikację obszarów wymagających dalszych badań i sugestie dotyczące potencjalnych ulepszeń w kontekście zastosowania technik analizy języka naturalnego w diagnozowaniu i leczeniu choroby Alzheimera.

Artykuł składa się z czterech części. Pierwsza skupia się na przykładach użycia technik przetwarzania języka naturalnego w diagnostyce i leczeniu choroby Alzheimera. W drugiej części omówione zostały konkretne usprawnienia i obszary, które wymagają większej uwagi w kontekście wykorzystania NLP. W kolejnym zostały wskazane kluczowe wyzwania związane z adaptacją NLP w chorobie Alzheimera. Ostatnia część to podsumowanie całego artykułu oraz wyszczególnienie najważniejszych wniosków.

## **2. Wykorzystanie NLP w diagnozowaniu oraz leczeniu choroby Alzheimera**

Przy wykorzystaniu technik przetwarzania języka naturalnego (NLP) w kontekście choroby Alzheimera pojawia się wiele możliwości zarówno diagnostycznych, jak i terapeutycznych. Zatem bardzo ważne jest zrozumienie ich potencjalnych korzyści i ograniczeń, co pomoże w identyfikacji najlepszych i najskuteczniejszych metod.

Jednymi z technik przetwarzania języka naturalnego, które są wykorzystywane w diagnostyce choroby Alzheimera, są automatyczne rozumienie mowy, wykrywanie zmian językowych czy wzorców językowych. Mimo tego iż wszystkie metody skupiają się na jak najwcześniejszym rozpoznaniu objawów tej choroby, nie każda jest równie skuteczna i praktyczna w różnych kontekstach klinicznych. Rozpoznanie mowy głównie opiera się na ręcznie przepisywanych transkrypcjach (Thomas i in., 2005; Habash i Guinn, 2012; Rentoumi i in., 2014), ale w ostatnich badaniach przedstawiono również:

- rozpoznawanie nagrań mowy do określenia obecności lub braku choroby (Zhou i in., 2016);
- nagrania video osób zdrowych i chorych (klinicznie zdiagnozowanych z chorobą Alzheimera), z których później wyekstrahowano nagrania audio do dalszej analizy (López de Ipiña i in., 2013);
- nagrania głosowe pacjentów pod kątem analizy markerów głosowych (König i in., 2015).

Jednakże badacze podkreślają, iż do wykorzystania analizy mowy w formie online lub do monitorowania w domu należy odpowiednio dostosować system, a nie skupiać się na ręcznej transkrypcji (Zhou i in., 2016). Z kolei wykrywanie wzorców

oraz zmian językowych dzięki przetwarzaniu języka naturalnego odbywa się w dużej mierze z tekstu mówionego (m.in. w badaniach: Fraser i in., 2016; König i in., 2015; López-de-Ipiña i in., 2013). Jedno badanie wykorzystało w tym celu konkretne cechy akustyczne wyodrębnione z mowy osób z łagodną postacią Alzheimerera (Meilán i in., 2014). Z kolei inne badanie skupiło się na predykcji choroby Alzheimerera na podstawie analizy mowy uzyskanej przy zadaniu opisu obrazów (Fraser i in., 2016). Z jednej strony jest to bardzo ważne, iż takie techniki istnieją i są w stanie pomóc we wczesnej diagnozie choroby Alzheimerera. Z drugiej strony należy poruszyć temat dostępności oraz praktyczności tych metod. Badania sugerują, że analiza mowy osób chorych na Alzheimerera może być skutecznym narzędziem diagnostycznym, ale wymaga dalszego opracowania w kontekście praktycznego zastosowania klinicznego. Wykrywanie wzorców i zmian językowych za pomocą NLP może być bardziej skomplikowane w zastosowaniu, ponieważ wymaga analizy dużych zbiorów danych tekstowych lub mówionych oraz opracowania zaawansowanych algorytmów do wykrywania subtelnych zmian. Kolejną kwestią, którą warto poruszyć w związku z tymi technikami przetwarzania języka naturalnego, jest ich skuteczność oraz precyzja. Badanie López de Ipiña i in. (2013) dotyczyło analizy cech mowy spontanicznej oraz reakcji emocjonalnych we wczesnych stadiach choroby Alzheimerera w zamyśle, aby opracować odpowiednie testy do wczesnej diagnozy tej choroby. Przetwarzanie języka naturalnego zostało wykorzystane do analizy tempa mowy, długości przerw oraz innych cech związanych z mową. W badaniu przeprowadzono wiele testów, jednak najbardziej interesujące wyniki (z perspektywy niniejszej pracy) wskazują, że system osiągał dobre rezultaty w określaniu najbardziej zaawansowanych stadiów choroby – umiarkowanej oraz zaawansowanej fazy choroby. Dla wczesnego stadium wyniki oscylowały w okolicach 60%. W przypadku klasyfikacji choroby występowały błędy dla osób znajdujących się między dwoma etapami choroby. Przejście od osoby zdrowej do wczesnego stadium, a czasami nawet umiarkowanej fazy choroby, może być trudne do wykrycia, jednakże nadal świadczy o postępie choroby. Z drugiej strony badanie Skirrow i in. (2022) skupiło się na walidacji zdalnego automatycznego zadania przypominania historii (ASRT) w celu oceny wczesnych zaburzeń poznawczych u osób starszych. Większość uczestników badania zgłosiła pozytywne doświadczenia związane z technicznymi aspektami aplikacji oraz łatwością obsługi, a także zainteresowaniem wykonywanymi zadaniami. Wyniki ogólnego dopasowania były niższe u uczestników z łagodnym upośledzeniem poznawczym lub łagodnym stadium choroby Alzheimerera, co sugeruje gorsze przypominanie historii w tej grupie. Wiarygodność wykorzystywanych historii w zadaniu ASRT była od umiarkowanej do silnej dla natychmiastowego i opóźnionego przypominania, co potwierdza spójność wyników z różnymi wersjami historii. Na podstawie tych danych można wywnioskować, iż testy te są dość niezawodne. Dodatkowo, ASRT wykazało umiarkowaną zgodność z ustalonymi testami poznawczymi, co sugeruje, że zadanie jest skuteczne w ocenie funkcji poznawczych. Ponadto badanie Meilán i in. (2014) skorzystało z odpowiednich cech akustycznych oraz tych związanych z mową

wśród osób z chorobą Alzheimera oraz zdrowych. W badaniu skupiano się na różnych miarach mowy i wykazano, iż na podstawie m.in. odsetka przerw między słowami, odsetka łamania głosu, drżenia głosu można określić z dokładnością do 84%, iż te cechy charakteryzują osoby z chorobą Alzheimera. Podobne wyniki uzyskało badanie de Lira i in. (2011), które wykorzystowało analizę mowy przez opisywanie obrazków. W tym badaniu osoby z chorobą Alzheimera uzyskiwały wyniki znacznie niższe niż grupa kontrolna w odniesieniu do analizy leksykalnej oraz składniowej. Zatem można powiedzieć, iż mowa osób z chorobą Alzheimera charakteryzowała się mniejszą złożonością zdań.

Dotychczasowe badania wskazują na dalszą potrzebę prowadzenia badań w obszarze technik przetwarzania języka naturalnego w diagnozowaniu choroby Alzheimera. Co więcej, dzięki zwiększeniu zainteresowania tym tematem oraz większej liczby prac badawczych można spekulować, iż poprawi to skuteczność algorytmów oraz precyzję diagnozy. Ponadto warto zwrócić również uwagę na możliwości połączenia standardowych metod diagnozowania z najnowszymi technikami NLP, co również przyczyni się do szybszego rozpoznania danego schorzenia.

### **3. Propozycje doskonalenia technik wykorzystania NLP w diagnozowaniu i leczeniu choroby Alzheimera**

Przeprowadzone analizy badań naukowych w obszarze technologii przetwarzania języka naturalnego wskazują, iż nawet diagnozowanie choroby Alzheimera standardowymi technikami, jak m.in. testami obrazowania mózgu, wykrywaniem odpowiednich biomarkerów, nie daje stu procentowej skuteczności (Shahidi i in., 2023 – radiomika + MRI; Engelborghs i in., 2008 – biomarker: tTau, pTau181). Zazwyczaj łączy się kilkanaście technik, aby w jak najmniejszym stopniu popełnić jakikolwiek błąd. Taki sam problem pojawia się we wdrożeniu technik przetwarzania języka naturalnego w diagnostyce choroby Alzheimera. Mimo tego iż większość z nich daje rezultaty na około 80% (m.in. Momota i in., 2023; König i in., 2015), nadal nie gwarantuje to odpowiednio wysokiej skuteczności diagnozowania. Jednakże warto zauważyć, iż automatyczne rozumienie mowy, wykrywanie zmian oraz wzorców językowych stanowią kluczowe elementy w identyfikacji objawów choroby Alzheimera. Bardzo ciekawym zagadnieniem jest fakt, iż niektóre techniki NLP nie są w stanie odróżnić różnych stadiów choroby Alzheimera (López de Ipiña i in., 2013), co wskazuje na to, iż większą uwagę należałoby skupić na personalizacji diagnozy. Niepewność w diagnostyce może wynikać z subtelnych różnic w języku i mowie pacjentów, które mogą być trudne do uchwycenia przez obecnie stosowane algorytmy. W związku z tym istnieje potrzeba skoncentrowania się na personalizacji diagnozy, aby efektywniej wyłapywać nawet te najmniejsze zmiany. Zatem w badaniach warto skupić się nie tylko na zmianach językowych, ale także unikalnych stylach mowy, które mogą nie tyle wskazywać na chorobę Alzheimera, ale braki językowe wynika-

jące z wady wymowy czy zaników obszarów w mózgu odpowiadających za mowę wraz ze starzeniem się.

Co więcej, skuteczność oraz praktyczność technik przetwarzania języka naturalnego w różnych kontekstach klinicznych stoją pod znakiem zapytania. Rozpoznawanie mowy oparte na ręcznej transkrypcji jest zbyt czasochłonne, uciążliwe oraz podatne na wiele błędów. Ponadto wykorzystanie analizy mowy w formie online lub jako domowe narzędzie wymaga specjalistycznego dostosowania. Z kolei wykrywanie wzorców oraz zmian językowych jest obiecującym zagadnieniem. Jednakże może być ono skomplikowane oraz wymaga znacznie większych zbiorów danych, co utrudnia praktyczne i efektywne wykorzystanie. Analiza tak dużych zbiorów danych może być czasochłonna i potrzebować zaawansowanych technik ich przetwarzania, aby wyodrębnić istotne wzorce i informacje. Oprócz tego istnieje konieczność uwzględnienia różnorodności danych oraz subtelnych różnic indywidualnych, co dodatkowo komplikuje proces analizy i interpretacji. Warto podkreślić, iż istnieje niewiele baz danych, które można wykorzystywać w trakcie diagnozy. Jedne z najbardziej znanych oraz rozpowszechnionych zbiorów danych dotyczących automatycznego wykrywania choroby Alzheimera opartych na spontanicznych rozmowach to, m.in.:

- DementiaBank (Boller i Becker, 2005) – największa udostępniona do analiz baza danych posiadająca 15 zbiorów danych w różnych językach, jak angielskim, mandaryńskim czy niemieckim. Nagrania audio są dostępne w przypadku każdego zbioru danych. Baza danych zawiera 241 próbek od zdrowych osób oraz 310 od osób ze zdiagnozowaną demencją w wieku od 45 do 90 lat (Qi i in., 2023).
- *The Carolinas Conversation Collection* (CCC) (Pope i Davis, 2011) – zbiór danych zawierający transkrypcje wideo oraz mowy osób w wieku powyżej 65. roku życia. Zawiera on około 200 rozmów z osobami z przewlekłymi chorobami oraz około 400 z chorującymi na Alzheimera (Qi i in., 2023).
- ADReSS (*The Alzheimer's Dementia Recognition through Spontaneous Speech*) – zbiór danych pochodzący z konferencji „Interspeech 2020” (Luz i in., 2020). Zawiera nagrania od 78 osób zdrowych oraz 78 nagrań od osób z chorobą Alzheimera w tym samym wieku oraz płci.

Zatem warto byłoby stworzyć zarówno globalne, jak i lokalne (w podziale na kraje) zbiory, a następnie porównać wyniki. Wymagałoby to ogromnych nakładów finansowych oraz czasu na przeprowadzenie i zebranie odpowiednich grup, jednakże warto rozważyć taką opcję. W kwestii praktyczności zastosowania technik przetwarzania języka naturalnego istnieją pewne ograniczenia i wątpliwości. Automatyczne rozpoznawanie mowy wymaga opracowania bardziej zaawansowanych algorytmów. Dodatkowo wykrywanie zmian oraz wzorców językowych powinno opierać się na dużo większych zbiorach danych, które dodatkowo umożliwią lepsze trenowanie i testowanie algorytmów, co przyczyni się do prawdopodobnie większej skuteczności tych technik przetwarzania języka naturalnego. Należy również wspomnieć, iż należy usprawnić i rozwinąć interfejsy użytkownika, które umożli-

wią łatwe i intuicyjne korzystanie z tych technik NLP przez np. personel medyczny. W związku z tym szpitale oraz kliniki medyczne mogłyby użytkować oprócz podstawowych metod diagnostycznych również techniki NLP, które służyłyby jako dodatkowe potwierdzenie diagnozy. Jednakże zanim to nastąpi, potrzeba odpowiednich procedur odnoszących się m.in. do wdrożenia, schematów postępowania, wykorzystania algorytmów, przechowywania danych. Co więcej, ważne jest stworzenie systemów automatycznej analizy mowy, które będą w czasie rzeczywistym zapewniać szczegółową, szybką i skuteczną diagnozę. Ponadto warto byłoby zainwestować w szkolenia dla personelu medycznego na temat wykorzystania nowych technik w praktyce klinicznej lub stworzyć odrębnych specjalistów, którzy wspomagaliby proces diagnozowania. Dzięki połączeniu wiedzy medycznej oraz technologicznej można byłoby skuteczniej przeprowadzać testy diagnostyczne.

Wobec tego należy również zwrócić uwagę na kwestię bezpieczeństwa danych oraz pacjentów, a także akceptacji wdrożenia technik NLP. Bardzo ważne w całym procesie będzie stworzenie oraz przestrzeganie wysokich standardów ochrony prywatności danych medycznych. Bezpieczeństwo przechowywania danych, ale także przetwarzanie zgodne z odpowiednimi regulacjami jak RODO (Rozporządzenie o Ochronie Danych Osobowych w państwach Unii Europejskiej) czy The Gramm Leach Bliley Act (w Stanach Zjednoczonych) (ICLG, 2024) będzie stanowiło jedno z podstawowych wyzwań tego przedsięwzięcia. Ponadto trzeba by było stworzyć grupę osób odpowiedzialnych za dostęp do danych oraz odpowiedzialnych za ich integrację, co jest kluczowe dla zachowania zaufania osób potencjalnie chorych/osób z chorobą Alzheimera do stosowania nowych technologii w obszarze opieki zdrowotnej. Z tym może wiązać się wprowadzenie szkoleń, informacji, narzędzi do edukacji dotyczących wykorzystania technik NLP w diagnozowaniu choroby Alzheimera. Przed wdrożeniem warto byłoby przeprowadzić badania ankietowe wyjaśniające potrzebę wdrożenia oraz doinformowania w tej kwestii respondentów – na początku grupę personelu medycznego, a dopiero później samych pacjentów oraz ich najbliższych. W ten sposób będzie można określić, jakie podejścia i emocje niesie ze sobą wprowadzenie tej zmiany. Wraz z tymi badaniami pojawia się kwestia etyki i bezpieczeństwa osób, które będą brały udział w diagnozowaniu choroby Alzheimera. Warto zwrócić uwagę, iż analiza danych nagrań głosowych powinna uwzględniać ochronę prywatności pacjentów. Wiąże się z tym wdrożenie odpowiednich zabezpieczeń, protokołów oraz poufność danych. Niestety istnieje ryzyko, iż osoby z chorobą Alzheimera z różnym stopniem zaawansowania mogą czuć dyskomfort czy niechęć do udziału w badaniach, zwłaszcza gdy ich nagrania głosowe będą wykorzystywane w celach diagnostycznych innych osób. Zatem należy odpowiednio zadbać o dane osoby przez wdrożenie konkretnych standardów opieki oraz należytego szacunku.

#### 4. Identyfikacja kluczowych wyzwań

W kwestii głębszego zrozumienia technik przetwarzania języka naturalnego w diagnozowaniu choroby Alzheimera oraz usprawnienia tego procesu bardzo ważne jest zrozumienie potencjalnych wyzwań i szans w opiece medycznej, które wynikają z postępów w przetwarzaniu języka naturalnego.

Zdrowie publiczne ma za zadanie dążenie do optymalnych wyników zdrowotnych w różnych populacjach poprzez opracowywanie i wdrażanie takich rozwiązań, które koncentrują się na możliwych do zmiany przyczynach złego stanu zdrowia (Jackson i Huston, 2016; Marmot i in., 2010; Arcaya i in., 2015). Osiąga się to przez efektywną diagnostykę chorób lub czynników ryzyka w populacji oraz danych populacji, zapobieganiu rozprzestrzenianiu chorób oraz oceny wyników tych działań (Ndumbe-Eyoh i in., 2016). Przyjmuje się taką koncepcję, według której podejmuje się odpowiednie decyzje mianowicie: PICO (*patient/ problem, intervention/ exposure, comparison/ outcome*), czyli pacjent/problem, interwencja/ ekspozycja, porównanie/ wynik, dzięki której jest możliwość odpowiedzi na dane pytania kliniczne lub te dotyczące zdrowia publicznego (Alonso-Coello i in., 2016).

Przetwarzanie języka naturalnego w obecnych czasach, gdzie informacje są publikowane i tworzone na dużą skalę, dało nowe możliwości badań opartych na tekście oraz podejmowaniu decyzji opartych na dowodach. W odniesieniu do opieki medycznej można wymienić następujące korzyści wynikające z wykorzystania NLP (Bačlic i in., 2020):

- szybka analiza dużej liczby nieustrukturyzowanych lub półstrukturyzowanych danych;
- możliwość wykorzystania NLP do identyfikacji populacji, interwencji i wyników potrzebnych do nadzorowania chorób, ich zapobiegania oraz promocji zdrowia;
- zdolność do przewidywania obecności depresji przed jej zarejestrowaniem w dokumentacji medycznej (Harris i in., 2014);
- możliwość przeprowadzania szybkiego przeglądu publikacji naukowych w celu uzyskania rekomendacji dotyczących zapobiegania lub leczenia chorób na podstawie najnowszych dowodów naukowych;
- wykorzystanie platform NLP do odpowiadania na pytania oraz tworzenia chatbotów, które mogą poprawić działania promocji zdrowia, angażując jednostki i dostarczając spersonalizowanego wsparcia czy porad.

Pomimo bardzo wielu zalet, które znacznie poprawiają efektywność opieki medycznej, począwszy od określenia potencjalnych czynników ryzyka chorób, wczesnej diagnozy, analizy danych aż do sugerowania, jakie leczenie w danym przypadku może być najbardziej skuteczne, przetwarzanie języka naturalnego niesie ze sobą też wiele wyzwań, które mogą znacznie ograniczyć możliwości wykorzystania. Głównymi problemami, z którymi należy się zmierzyć, są:

- zależność od dostępności oraz jakości danych treningowych dla danych modeli (Tatman i Conner, 2017). Chociaż w ostatnich latach wzrasta liczba bezpłatnie



dostępnych zbiorów danych biomedycznych i wstępnie przeszkolonych modeli, dostępność tych dotyczących koncepcji zdrowia publicznego pozostaje ograniczona;

- konieczność eliminacji stronniczości w danych – nieuwzględnienie stronniczości w opracowywaniu (np. adnotacji danych), wdrażaniu (np. wykorzystanie wstępnie przeszkolonych platform) i ocenie modeli NLP mogłoby zagrozić osiągnięciom modelu i pogłębić istniejące nierówności zdrowotne (Baclic i in., 2020). Mimo to należy również zwracać uwagę na inne czynniki, które model musi uwzględniać na przykład: grupy wiekowe, status społeczno-ekonomiczny w trakcie analizy danych medycznych przez media społecznościowe;
- ograniczony dostęp do danych – np. w Kanadzie dane zdrowotne z powodu regulacji bezpieczeństwa oraz poufności niechętnie udostępnia się możliwość ich przeglądania. Co ciekawe badanie Forda i in. (2019) wykazało, iż większość użytkowników mediów społecznościowych nie wyraża zgody na analizę swoich danych w celu identyfikacji problemów związanych ze zdrowiem psychicznym;
- ograniczone zestawy danych – jest zbyt mało zestawów danych, które mogłyby być użyte w badaniach z wykorzystaniem różnych technik NLP;
- ocena i ewaluacja modeli NLP – w celu zapewnienia zgodności z normami etycznymi oraz poprawności działania modeli należy je odpowiednio analizować. Nawet jeśli modele osiągają wysokie wyniki, nie oznacza to koniecznie, że w pełni rozumieją język. Warto jednak pamiętać, że modele mogą być nadal skuteczne w ekstrakcji informacji, klasyfikacji czy choćby zadań związanych z przewidywaniem, nawet jeśli nie rozumieją języka w pełni.

Techniki przetwarzania języka naturalnego (NLP) oferują obiecujące korzyści w opiece medycznej, jednakże istnieją również wyzwania związane z NLP, w tym ograniczony dostęp do odpowiednio skategoryzowanych i oznakowanych danych treningowych, konieczność eliminacji stronniczości w danych, czemu należałoby się przyjrzeć w dalszych badaniach w tym temacie.

## 5. Zakończenie

Analiza porównawcza dotychczas wykonanych badań oraz przegląd literatury w zakresie omawianej dziedziny pozwoliły na wskazanie kilku obszarów, w których NLP zarówno wykazuje się istotną rolą w danym procesie, ale również aspekty, które mogą mieć istotny wpływ na poprawę nie tylko diagnostyki choroby Alzheimera, ale także poprawę opieki nad osobami cierpiącymi na to schorzenie. Zidentyfikowano potencjalne wyzwania, z którymi muszą zmierzyć się badacze, jak implementacja oraz dalszy rozwój technologii przetwarzania języka naturalnego pod względem w szczególności odpowiedniej jakości danych treningowych i regulacji dotyczących prywatności danych pacjentów.

Główne wnioski płynące z tego artykułu to:

- konieczność dalszych badań i rozwój technologii przetwarzania języka naturalnego do skuteczniejszego wykorzystania w praktyce klinicznej;
- wykorzystanie w NLP bardziej zaawansowanych modeli, np. głębokiego uczenia, co może przyczynić się do usprawnienia procesu diagnozowania choroby Alzheimera, personalizacji leczenia w tym dostosowywania odpowiednich terapii oraz monitorowania ich postępów;
- zwiększona potrzeba uwzględnienia personalizacji diagnozy, gdyż objawy choroby Alzheimera mogą się różnić między poszczególnymi przypadkami. Wynika to zarówno z czynników, takich jak wiek, przebyte choroby, urazy mózgu, genetyki, czynników środowiskowych oraz indywidualnych;
- stworzenie globalnych oraz krajowych baz danych z różnorodnymi aspektami danej choroby – preferowane byłyby nagrania wideo, z których można by pobrać dane audio, ale również odczytać mimikę i zachowanie danych osób, które na pewnym etapie mogą być indykatorem postępującego schorzenia. Ponadto warto byłoby uwzględnić możliwość aktualizacji tych danych, np. w których przypadkach i jakie objawy doprowadziły do diagnozy;
- uwzględnienie w dalszych badaniach oraz przy tworzeniu bazy danych zróżnicowania populacji osób z chorobą Alzheimera oraz specyfiki objawów tego schorzenia, co pozwoli na porównywanie danych między populacjami;
- konieczność dalszej kontynuacji interdyscyplinarnych badań i współpracy – nauki medyczne, nauki technologiczne, nauki o języku. Poprzez zwiększenie wysiłków naukowych z różnych dziedzin można będzie maksymalnie wykorzystać potencjał w diagnozowaniu, leczeniu oraz poprawie opieki nad osobami cierpiącymi na to schorzenie neurodegeneracyjne;
- warto rozważyć współpracę pomiędzy środowiskiem badań klinicznych, sektorem naukowym oraz technologicznym, aby przyspieszyć innowacyjne rozwiązania i wdrożyć je szybciej w trakcie, gdy będą dopracowywane szczegóły technologii przetwarzania języka naturalnego w praktyce;
- warto również podkreślić znaczenie edukacji i szkoleń dla personelu medycznego zakres oraz znaczenie technologii przetwarzania języka naturalnego przy schorzeniu Alzheimera. Świadomość możliwości, jakie niesie ze sobą ta technologia nie tylko w diagnostyce, może prowadzić do lepszej integracji w codziennej praktyce klinicznej;
- dodatkowe inwestycje w rozwój tej technologii mogą przynieść korzyści nie tylko dla pacjentów i personelu medycznego, ale również dla systemów opieki zdrowotnej. Dlatego warto pogłębiać wysiłki badawcze i innowacyjne w walce z tym rozprzestrzeniającym się schorzeniem neurodegeneracyjnym.

Dodatkowo, technologie NLP wykazują obiecujący potencjał nie tylko w diagnostyce, ale także w leczeniu choroby Alzheimera. Identyfikacja obszarów z największym potencjałem dla NLP w leczeniu choroby Alzheimera jest kluczowym krokiem w dalszym rozwoju tej dziedziny. W związku z tym istnieje potrzeba dalszego

skupienia się na badaniach nad zastosowaniem NLP w leczeniu oraz w wsparciu opiekunów pacjentów, co może przyczynić się do poprawy jakości życia zarówno chorych, jak i ich opiekunów. NLP może być używane do monitorowania i analizy komunikacji pacjentów w celu identyfikacji subtelnych zmian w języku, które mogą wskazywać na progresję choroby. Zastosowanie NLP w terapii może umożliwić personalizację leczenia oraz efektywne monitorowanie postępów, a także identyfikację najbardziej skutecznych metod terapeutycznych. Analiza komunikacji i potrzeb opiekunów może dostarczyć cennych informacji do poprawy systemów wsparcia oraz wprowadzenia nowych procedur opieki. NLP może również wspierać edukację i profilaktykę choroby Alzheimera, zwiększając świadomość społeczeństwa, wspierając wcześniejsze wykrywanie oraz wdrażanie leczenia.

## Literatura

- Alonso-Coello, P. i in. (2016). GRADE Evidence to Decision (EtD) Frameworks: A Systematic and Transparent Approach to Making Well Informed Healthcare Choices. 1: Introduction. *BMJ (Clinical research ed.)*, (353), i2016. <https://doi.org/10.1136/bmj.i2016>
- Arcaya, M. C., Arcaya, A. L. i Subramanian S. V. (2015). Inequalities in Health: Definitions, Concepts, and Theories. *Global Health Action*, (8), 27106. <https://doi.org/10.3402/gha.v8.27106>
- Baclic, O., Tunis, M., Young, K., Doan, C., Swerdfeger, H., i Schonfeld, J. (2020). Challenges and Opportunities for Public Health Made Possible by Advances in Natural Language Processing. *Canada Communicable Disease Report = Releve des Maladies Transmissibles au Canada*, 46(6), 161-168. <https://doi.org/10.14745/ccdr.v46i06a02>
- Boller, F., i Becker, J. (2005). *Dementiabank Database Guide*. University of Pittsburgh.
- de Lira, J. O., Ortiz, K. Z., Campanha, A. C., Bertolucci, P. H., i Minett, T. S. (2011). Microlinguistic Aspects of the Oral Narrative in Patients with Alzheimer's Disease. *International psychogeriatrics*, 23(3), 404-412. <https://doi.org/10.1017/S1041610210001092>
- Dubois, B., i in. (2016). Preclinical Alzheimer's Disease: Definition, Natural History, and Diagnostic Criteria. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 12(3), 292-323. <https://doi.org/10.1016/j.jalz.2016.02.002>
- Engelborghs, S., i in. (2008). Diagnostic Performance of a CSF-biomarker Panel in Autopsy-Confirmed Dementia. *Neurobiology of Aging*, 29(8), 1143-1159. <https://doi.org/10.1016/j.neurobiolaging.2007.02.016>
- Erkinen, M. G., Kim, M. O., i Geschwind, M. D. (2018). Clinical Neurology and Epidemiology of the Major Neurodegenerative Diseases. *Cold Spring Harbor Perspectives in Biology*, 10(4), 1-44. <https://doi.org/10.1101/cshperspect.a033118>
- Ford, E., Curlew, K., Wongkoblap, A., i Curcin, V. (2019). Public Opinions on Using Social Media Content to Identify Users with Depression and Target Mental Health Care Advertising: Mixed Methods Survey. *JMIR Mental Health*, 6(11), e12942. <https://doi.org/10.2196/12942>
- Fraser, K. C., Meltzer, J. A., i Rudzicz, F. (2016). Linguistic Features Identify Alzheimer's Disease in Narrative Speech. *Journal of Alzheimer's disease: JAD*, 49(2), 407-422. <https://doi.org/10.3233/JAD-150520>
- Gagliardi, G., Kokkinakis, D., i Duñabeitia, J. A. (2021). Editorial: Digital Linguistic Biomarkers: Beyond Paper and Pencil Tests. *Frontiers in Psychology*, (12), 752238. <https://doi.org/10.3389/fpsyg.2021.752238>

- Gaweł, M. i Potulska-Chromik, A. (2015). Choroby eurodegeneracyjne: choroba Alzheimera i Parkinsona. *Postępy Nauk Medycznych*, 28(7), 468-476.
- Habash, A., i Guinn, C. (2012). *Language Analysis of Speakers with Dementia of the Alzheimer's Type*. In Association for the Advancement of Artificial Intelligence (AAAI) Fall Symposium, 2012, 8-13.
- Harris, J. K., Mansour, R., Choucair, B., Olson, J., Nissen, C., i Bhatt, J. (2014). Health Department Use of Social Media to Identify Foodborne Illness – Chicago, Illinois, 2013-2014. *MMWR. Morbidity and Mortality Weekly Report*, 63(32), 681-685.
- ICLG. (2024). *Data Protection Laws and Regulations USA 2023-2024*. Pobrane 4 maja 2024 z <https://iclg.com/practice-areas/data-protection-laws-and-regulations/usa>
- Jackson, B., i Huston, P. (2016). Advancing Health Equity to Improve Health: The Time is Now. *Health Promotion and Chronic Disease Prevention in Canada: Research, Policy and Practice*, 36(2), 17-20. <https://doi.org/10.24095/hpcdp.36.2.01>
- König, A., i in. (2015). Automatic Speech Analysis for the Assessment of Patients with Predementia and Alzheimer's Disease. *Alzheimer's & Dementia (Amsterdam, Netherlands)*, 1(1), 112-124. <https://doi.org/10.1016/j.dadm.2014.11.012>
- López-de-Ipiña, K., i in. (2013). On the Selection of Non-Invasive Methods Based on Speech Analysis Oriented to Automatic Alzheimer Disease Diagnosis. *Sensors*, 13(5), 6730-6745. <https://doi.org/10.3390/s130506730>
- Luz, S., Haider, F., de la Fuente, S., Fromm, D., i MacWhinney, B. (2020). Alzheimer's Dementia Recognition through Spontaneous Speech: The Address Challenge. W: *Proceedings of Interspeech 2020* (Grenoble: ISCA), 2172-2176.
- Marmot, M., Allen, J., Goldblatt, P., Boyce, T., McNeish, D., Grady, M., i Geddes, I. (2010). *The Marmot Review. Fair Society, Healthy Lives: Strategic Review of Health Inequalities in England post-2010*. UCL Institute of Health Equity. Pobrane 6 kwietnia 2024 z <http://www.parliament.uk/documents/fair-society-healthy-lives-full-report.pdf>
- Meilán J. J. G., Martínez-Sánchez F., Carro J., López D., Millian-Morell L., i Arana J. (2014). Speech in Alzheimer's Disease: Can Temporal and Acoustic Parameters Discriminate Dementia? *Dementia and Geriatric Cognitive Disorders*, 37(5-6), 327-334. <https://doi.org/10.1159/000356726>
- Momota, Y., i in. (2023). Language Patterns in Japanese Patients with Alzheimer Disease: A Machine Learning Approach. *Psychiatry and Clinical Neurosciences*, 77(5), 273-281. <https://doi.org/10.1111/pcn.13526>
- Myszczynska, M. A., i in. (2020). Applications of Machine Learning to Diagnosis and Treatment of Neurodegenerative Diseases. *Nature Reviews Neurology*, 16(8), 440-456. <https://doi.org/10.1038/s41582-020-0377-8>
- National Cancer Institute. (b.d.). *Neurodegenerative Disorder*. Cancer.gov. Pobrano 7 maja 2024 z <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/neurodegenerative-disorder>
- Ndumbe-Eyoh, S., Dyck, L., i Clement, C. (2016). *Common Agenda for Public Health Action on Health Equity*. National Collaborating Centre for Determinants of Health, St Francis Xavier University.
- Pope, C., i Davis, B. (2011). Finding a Balance: The Carolinas Conversation Collection. *Corpus Linguistics and Linguistic Theory*, 7(1), 143-161. <https://doi.org/10.1515/cllt.2011.007>
- Poradnik Gemini. (b.d.). *Choroby neurologiczne – czym są? Rodzaje, objawy, leczenie, profilaktyka*. Gemini.pl. Pobrano 7 maja 2024 z <https://gemini.pl/poradnik/zdrowie/choroby-neurologiczne/>
- Qi, X., Zhou, Q., Dong, J., i Bao, W. (2023). *Noninvasive Automatic Detection of Alzheimer's Disease from Spontaneous Speech: A Review*. *Frontiers in Aging Neuroscience*, (15). <https://doi.org/10.3389/fnagi.2023.1224723>
- Rentoumi, V., Raoufian, L., Ahmed, S., de Jager, C. A., i Garrard, P. (2014). Features and Machine Learning Classification of Connected Speech Samples from Patients with Autopsy Proven Alzheimer's Disease with and without Additional Vascular Pathology. *Journal of Alzheimer's Disease: JAD*, 42 Suppl 3, S3-S17. <https://doi.org/10.3233/JAD-140555>

- Shahidi, R., i in. (2023). Diagnostic Performance of MRI Radiomics for Classification of Alzheimer's Disease, Mild Cognitive Impairment, and Normal Subjects: A Systematic Review and meta-Analysis. *Aging Clinical and Experimental Research*, 35(11), 2333-2348. <https://doi.org/10.1007/s40520-023-02565-x>
- Shusharina, N., i in. (2023). Modern Methods of Diagnostics and Treatment of Neurodegenerative Diseases and Depression. *Diagnostics*, 13(3), 1-27. <https://doi.org/10.3390/diagnostics13030573>
- Skirrow, C., i in. (2022). Validation of a Remote and Fully Automated Story Recall Task to Assess for Early Cognitive Impairment in Older Adults: Longitudinal Case-Control Observational Study. *JMIR aging*, 5(3). <https://doi.org/10.2196/37090>
- Tagaris, A., Kollias, D., Stafylopatis, A., Tagaris, G., i Kollias, S. (2018). Machine Learning for Neurodegenerative Disorder Diagnosis—Survey of Practices and Launch of Benchmark Dataset. *International Journal of Artificial Intelligence Tools*, 27(3), 1-27. <https://doi.org/10.1142/S0218213018500112>
- Tatman, R., i Conner, K. (2017). Effects of Talker Dialect, Gender & Race on Accuracy of Bing Speech and YouTube Automatic Captions. *Proc Interspeech*, 934-8. <https://doi.org/10.21437/Interspeech.2017-1746>
- Thau, L., Reddy, V., i Singh, P. (2022). Anatomy, Central Nervous System. In *StatPearls*. StatPearls Publishing. Pobrano 7 maja 2024 r z <https://www.ncbi.nlm.nih.gov/books/NBK542179/>
- Thomas, C., Keselj, V., Cercone, N., Rockwood, K., i Asp, E. (2005). *Automatic Detection and Rating of Dementia of Alzheimer Type Through Lexical Analysis of Spontaneous Speech*. IEEE International Conference Mechatronics and Automation, 1569-1574. <https://doi.org/10.1109/ICMA.2005.1626789>
- Wu, Y., Le, W., i Jankovic, J. (2011). Preclinical Biomarkers of Parkinson Disease. *Archives of Neurology*, (68), 22-30. <https://doi.org/10.1001/archneurol.2010.321>
- Zhou, L., Fraser, K. C., i Rudzicz, F. (2016). *Speech Recognition in Alzheimer's Disease and in its Assessment*. Proceeding of Interspeech 2016, 1948-1952. <https://doi.org/10.21437/Interspeech.2016-1228>

## Application of Natural Language Processing (NLP) in the Diagnosis and Treatment of Alzheimer's Disease

**Abstract:** The following article addresses the role of various natural language processing (NLP) techniques in the diagnosis and treatment of Alzheimer's disease. Additionally, the paper examines the interactions between linguistics and neurobiology in the context of these techniques. The work is based on sources such as scientific literature, specialist journals, and information available on the Internet. The research methods used in this thesis include a literature review, a comparative analysis of specific studies, and the identification of areas for further research along with proposals for improving the application of natural language processing techniques in Alzheimer's disease. The comparative analysis of studies revealed the necessity for further research, the development of NLP with the use of deep learning, personalised treatment, the creation of global databases, and interdisciplinary collaboration to enhance the benefits for patients, medical personnel, and healthcare systems.

**Keywords:** natural language processing, NLP, Alzheimer's disease