

A G H

AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

**WYDZIAŁ ELEKTROTECHNIKI, AUTOMATYKI,
INFORMATYKI I INŻYNIERII BIOMEDYCZNEJ**

Praca dyplomowa

*Detekcja anomalii w sygnale mowy u osób z chorobą Parkinsona
Anomalies detection in speech signals in people with Parkinson's
disease*

Autor: *Małgorzata Szwed*
Kierunek studiów: *Inżynieria Biomedyczna*
Opiekun pracy: *dr inż. Daria Hemmerling*

Kraków, 2023

Uprzedzony o odpowiedzialności karnej na podstawie art. 115 ust. 1 i 2 ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (t.j. Dz.U. z 2006 r. Nr 90, poz. 631 z późn. zm.): „Kto przywłaszcza sobie autorstwo albo wprowadza w błąd co do autorstwa całości lub części cudzego utworu albo artystycznego wykonania, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do lat 3. Tej samej karze podlega, kto rozpościera bez podania nazwiska lub pseudonimu twórcy cudzy utwór w wersji oryginalnej albo w postaci opracowania, artystycznego wykonania albo publicznie zniekształca taki utwór, artystyczne wykonanie, fonogram, videogram lub nadanie.”, a także uprzedzony o odpowiedzialności dyscyplinarnej na podstawie art. 211 ust. 1 ustawy z dnia 27 lipca 2005 r. Prawo o szkolnictwie wyższym (t.j. Dz. U. z 2012 r. poz. 572, z późn. zm.): „Za naruszenie przepisów obowiązujących w uczelni oraz za czyny uchybiające godności studenta student ponosi odpowiedzialność dyscyplinarną przed komisją dyscyplinarną albo przed sądem koleżeńskim samorządu studenckiego, zwanym dalej «sądem koleżeńskim».”, oświadczam, że niniejszą pracę dyplomową wykonałem(-am) osobiście i samodzielnie i że nie korzystałem(-am) ze źródeł innych niż wymienione w pracy.

Serdecznie dziękuję . . . tu ciąg dalszych podziękowań np. dla promotora, żony, sąsiada itp.

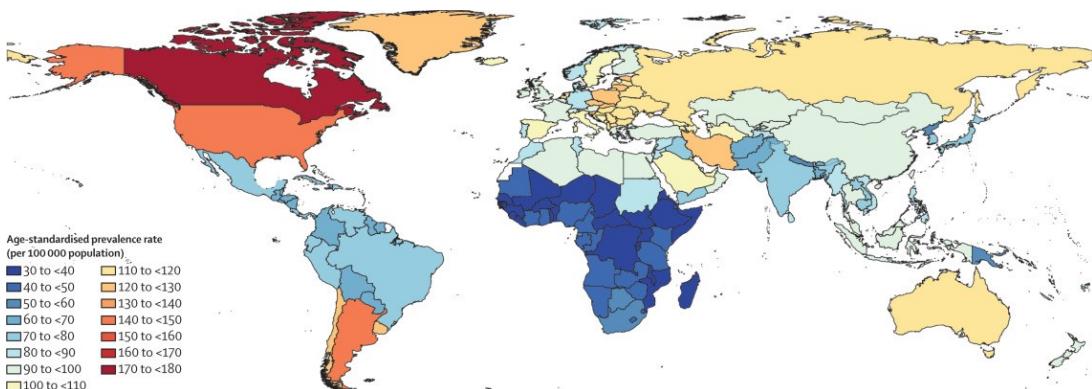
Spis treści

1. Wprowadzenie	7
1.1. Cel pracy	9
1.2. Zakres pracy	9
2. Problematyka zagadnienia	11
2.1. Znaczenie głosu w chorobie Parkinsona	11
2.2. Metody diagnozowania i monitorowania	14
2.3. Terapia osób chorych.....	16
3. Analiza rozwiązań do automatycznej diagnostyki choroby Parkinsona	19
3.1. Rodzaj wykorzystywanych danych	19
3.2. Metody klasyfikacji	21
3.3. Wyzwania związane z systemami automatycznej diagnostyki.....	25
4. Materiał i metoda badawcza	31
4.1. Materiał badawczy.....	31
4.2. Parametryzacja sygnału akustycznego	36
4.3. Metody klasyfikacji	40
4.4. Metody ewaluacji wyników.....	43
5. Wyniki badań	47
5.1. Samogłoska /a/.....	47
5.2. Samogłoska /e/.....	48
5.3. Samogłoska /i/	48
5.4. Samogłoska /o/	48
5.5. Samogłoska /u/	48
5.6. Połączenie samogłosek /a/, /e/, /i/, /o/, /u/	48
5.7. Zbiorcze podsumowanie wyników	48
6. Analiza i interpretacja wyników.....	51
6.1. Samogłoski a informacja diagnostyczna	51
6.2. Architektury CNN a skuteczność klasyfikacji.....	51

7. Podsumowanie	53
Bibliografia	55

1. Wprowadzenie

Choroba Parkinsona (ang. *Parkinson Disease, PD*) to zwyrodnieniowe schorzenie mózgu, które wiąże się z objawami ruchowymi, takimi jak spowolnienie ruchowe, drżenie, sztywność oraz zaburzenia chodu i równowagi. Ponadto może prowadzić do różnorodnych powikłań niemotorycznych, obejmujących zaburzenia poznawcze, stany psychiczne, trudności ze snem oraz dolegliwości sensoryczne, w tym ból. Początkowe objawy często rozwijają się stopniowo, nasilając się w miarę upływu czasu. Postęp choroby prowadzi do znacznego stopnia niepełnosprawności, co może wymagać wsparcia i opieki. U wielu osób zdiagnozowanych z chorobą Parkinsona występują także zmiany w sferze psychicznej i behawioralnej, takie jak trudności ze snem, depresja, problemy z pamięcią oraz uczucie przewlekłego zmęczenia.



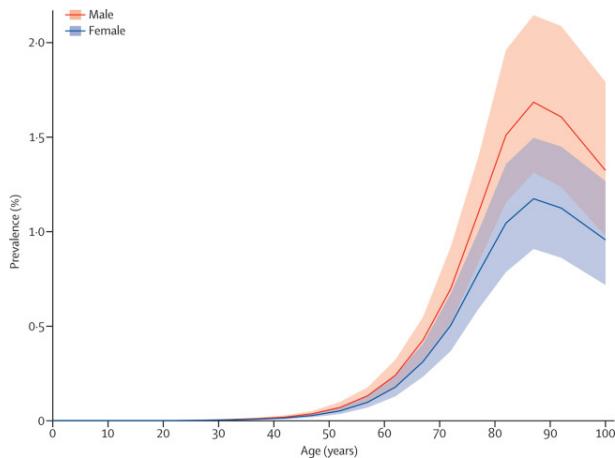
Rys. 1.1. Choroba Parkinsona na świecie [1]

Zgodnie z danymi przedstawionymi w raporcie Światowej Organizacji Zdrowia [2], choroba Parkinsona stanowi obecnie narastający problem na skalę światową. Zarówno wskaźniki niepełnosprawności, jak i zgony związane z tą chorobą rosną szybciej niż w przypadku innych zaburzeń neurologicznych.

W ciągu ostatnich 25 lat zaobserwowano podwojenie częstości występowania PD na całym świecie. Globalne szacunki na rok 2019 wskazują, że liczba osób cierpiących na PD przekroczyła 8,5 miliona, co oznacza wzrost o 81% w porównaniu z danymi z roku 2000. Jednocześnie liczba zgonów związanych z PD wyniosła 329 000, co stanowi wzrost o ponad 100% w porównaniu z rokiem 2000[1]. W Polsce z chorobą Parkinsona zmaga się około 100 tys. pacjentów, z czego około 20% jest już w stadium zaawansowanym według informacji przekazywanych przez Fundację Chorób Mózgu. Ponadto co roku w naszym kraju wykrywanych jest ok. 8 tys. nowych zachorowań.

PD jest istotną sprawą dotyczącą zdrowia publicznego, ponieważ jej częstotliwość występowania związana jest ze zjawiskiem starzającym się społeczeństwa (Rys. 1.1). Razem z innymi chorobami neurodegeneracyjnymi, takimi jak choroba Alzheimera, ma szanse stać się drugą zaraz za nowotworami przyczyną zgonów do 2040 roku (WHO).

Przyczyna PD nie jest znana, ale uważa się, że powstaje w wyniku złożonej interakcji pomiędzy czynnikami genetycznymi i narażeniem na czynniki środowiskowe, takie jak pestycydy, rozpuszczalniki i zanieczyszczenia powietrza. Niektóre przypadki wydają się dziedziczne, a kilka można przypisać określonym wariantom genetycznym. Chociaż uważa się, że genetyka odgrywa rolę w chorobie Parkinsona, w większości przypadków choroba nie występuje rodzinie[3].



Rys. 1.2. Rozpowszechnienie choroby Parkinsona w zależności od wieku [1]

Mimo że każdy może być narażony na ryzyko rozwoju tej choroby, to częściej występuje ona u mężczyzn niż u kobiet, a wiek stanowi kluczowy element wpływający na ryzyko zachorowania, co można zobaczyć na Rys.1.2. Statystyki pokazują, że ryzyko zachorowania rośnie wraz z wiekiem, chociaż choroba może dotyczyć także młodszych osób (nawet w wieku 20 lat). U większości pacjentów po raz pierwszy choroba rozwija się po 60 roku życia, około 5% do 10% doświadcza jej początku przed 50 rokiem życia. Postacie choroby Parkinsona o wczesnym początku są często, choć nie zawsze, dziedziczne i niektóre formy zostały powiązane z określonymi zmianami w genach[3].

Proces diagnozowania choroby jest niezwykle złożony i czasochłonny, nie istnieje obecnie pojedyncze badanie pozwalające na postawienie diagnozy. W związku z tym poszukuje się nowych rozwiązań, które mogłyby usprawnić ten proces. Coraz częściej wykorzystuje się metody uczenia maszynowego i sztucznej inteligencji w dziedzinie medycyny. W niniejszej pracy analizowany jest aktualny stan rzeczy oraz przedstawiane jest proponowane rozwiązanie, dotyczące automatycznej diagnostyki choroby Parkinsona na podstawie głosu.

1.1. Cel pracy

Celem pracy jest detekcja choroby Parkinsona na podstawie sygnału głosowego z wykorzystaniem metod uczenia maszynowego. Obejmuje to dokładny przegląd literaturowy ze szczególnym uwzględnieniem aktualnie najlepszych algorytmów dostępnych w literaturze. Pośrednim celem jest ocena skuteczności wybranych architektur konwolucyjnych sieci neuronowych (CNN) oraz analiza, która z wypowiadanych przez pacjentów samogłosek niesie ze sobą najwięcej informacji diagnostycznej.

1.2. Zakres pracy

2. Problematyka zagadnienia

2.1. Znaczenie głosu w chorobie Parkinsona

Najbardziej widoczne oznaki i objawy choroby Parkinsona pojawiają się, gdy komórki nerwowe w zwojach podstawy mózgu, obszarze kontrolującym ruch, ulegają uszkodzeniu lub obumierają. Zwykle te komórki nerwowe lub neurony wytwarzają dopaminę. Kiedy neurony obumierają lub ulegają uszkodzeniu, wytwarzają mniej dopaminy, co powoduje problemy z poruszaniem się związane z chorobą. Na ten moment nie wiadomo co powoduje śmierć neuronów. Zanikają również zakończenia nerwowe, które wytwarzają norepinefrynę, główny przekaźnik chemiczny współczulnego układu nerwowego, który kontuluje wiele funkcji organizmu, takich jak tężno i ciśnienie krwi. Utrata norepinefryny może pomóc wyjaśnić niektóre cechy choroby Parkinsona związane z brakiem ruchu, takie jak zmęczenie, nieregularne ciśnienie krwi, zmniejszony ruch pokarmu w przewodzie pokarmowym i nagły spadek ciśnienia krwi, gdy osoba wstaje z pozycji siedzącej lub leżącej.

Do czterech głównych objawów choroby Parkinsona zalicza się:

- drżenie rąk, ramion, nóg, szczęki lub głowy,
- sztywność mięśni, gdy mięśnie pozostają skurzone przez długi czas,
- powolność ruchu,
- zaburzenia równowagi i koordynacji, czasami prowadzące do upadków.

Zarówno objawy choroby Parkinsona, jak i tempo jej postępu mogą znacznie różnić się wśród poszczególnych osób. Na wczesnym etapie choroby objawy są subtelne i kształtują się stopniowo. Często zaczynają się od jednej strony ciała lub nawet jednej kończyny. Niektórzy pacjenci doświadczają pewnych zwiastunów przed wystąpieniem charakterystycznych cech, takich jak sztywność czy drżenie. Mogą to być trudności ze snem, problemy z wypróżnianiem, utrata węchu czy także zespół niespokojnych nóg. Warto jednak zaznaczyć, że niektóre z wymienionych objawów mogą również występować w procesie naturalnego starzenia się [3].

Chociaż tempo postępu choroby Parkinsona zazwyczaj jest powolne, w końcu ma to wpływ na codzienne funkcjonowanie osoby dotkniętej tą dolegliwością. Wykonywanie zwykłych czynności, takich jak praca, prowadzenie domu czy uczestnictwo w spotkaniach towarzyskich z przyjaciółmi, może stawać się wyzwaniem.

Choroba Parkinsona jest spowodowana zaburzeniem funkcji układu nerwowego, co prowadzi do występowania objawów w różnych obszarach organizmu, w tym również w sferze głosu. Objawy zaburzeń mowy dotyczą między 60 a 80% pacjentów, a powodowane są przede wszystkim przez deficit czynnościowy krtani, zmniejszoną pojemność życiową płuc, osłabioną pracę mięśni mimicznych oraz zmniejszony napęd mówienia [4]. Zwykle stają się widoczne w średniozaawansowanym stadium schorzenia, co oznacza, że przez długi okres mowa pozostaje relatywnie nienaruszona. Rozpoznanie tych zaburzeń może być niekiedy trudne, gdyż wymaga odróżnienia, czy powstały one na skutek samej choroby, czy też są rezultatem naturalnego procesu starzenia się organizmu.

Objawy zaburzeń mowy i głosu związane z chorobą Parkinsona nie są łatwo dostrzegalne dla osób bez specjalistycznej wiedzy w tej dziedzinie. Przeważnie zdolność rozumienia mowy pozostaje niezmieniona. Niemniej jednak, w trakcie spontanicznych wypowiedzi pacjenci zaczynają ograniczać ilość przekazywanych informacji i mogą napotykać trudności w składaniu pełnych zdań. Trudności te nie wynikają koniecznie z ubytku słownictwa, ale raczej z nieprawidłowego doboru słów. Wskazujące na podłożę chorobowe objawy obejmują między innymi [5, 6]:

- mowę powolną, monotonną i przerywaną,
- nadmierne ślinienie się,
- niewyraźną i zamazaną artykulację,
- skrócony czas fonacji
- chuchający i tremolujący głos
- spłaszczoną barwę i obniżone natężenie
- niewłaściwą koordynację mięśni nasady, które mogą być zwiotczałe lub zbyt napięte,
- czasami przyspieszenie tempa wypowiedzi w jej końcowej fazie, co może utrudnić zrozumienie pacjenta.

Wymienione objawy pojawiają się w średniozaawansowanym stadium choroby i są wystarczająco wyraźne, aby mogły zostać zauważone słuchowo przez specjalistów. Niemniej jednak badania wskazują, że istnieją subtelne zmiany w głosie, które pojawiają się jeszcze wcześniej, nawet w fazie przedobjawowej [7].

W roku 2000 przeprowadzono badanie akustyczne i percepcyjne cech głosu pacjentów z chorobą Parkinsona, zależnie od stopnia nasilenia choroby [8]. W nagraniach głosowych, składających się z przedłużonej samogłoski /a/, śpiewu gamy oraz 1-minutowego monologu, stwierdzono, że głosy pacjentów z PD, zarówno we wczesnych, jak i późniejszych stadiach choroby, charakteryzowały się ograniczoną percepcyjnie zmiennością tonu i głośności, chropowatością oraz zmniejszoną głośnością.

Badanie sugerowało również, że głosy pacjentów z PD wykazywały nadmierne drganie, wysoką częstotliwość podstawową (szczególnie u mężczyzn) oraz zmniejszoną zmienność częstotliwości podstawowej (szczególnie u kobiet). Część z tych cech głosu nie wydawała się pogarszać w miarę postępu

choroby, jednak cechy takie jak oddech, monotonność i jednolitość mowy, niska głośność oraz ograniczony maksymalny zakres częstotliwości fonacyjnej były bardziej zauważalne w późniejszych stadiach choroby Parkinsona.

Podobne badanie opisane w publikacji [9] wykazało, że w porównaniu z grupą kontrolną, pacjenci z PD wykazywali wyższy jitter, niższy stosunek harmonicznych do szumów (H/N), mniejszą zmienność częstotliwości i intensywności mowy oraz niższy zakres fonacyjny, oraz wyższą częstotliwość obecności głosu o niskim natężeniu. Wskazano również, że te cechy nie wykazywały znaczącego związku z czasem trwania choroby.

Mnogość objawów, które są zauważalne w głosie, motywuje do uwzględnienia ich w diagnostyce. Prowadzone są badania, które wykorzystują analizę mowy do wykrywania patologii i schorzeń związanych z narządem głosu, takich jak ostre zapalenie krtani czy porażenie nerwu krtaniowego wstecznego. Może to w przyszłości pozwolić na identyfikację problemów zdrowotnych, na przykład wśród osób pracujących głosem, jak nauczyciele, bez konieczności inwazyjnych badań gardła. Podobne podejście można zastosować do diagnozowania i monitorowania chorób neurodegeneracyjnych. Badania naukowe wskazują, że analiza głosu może stanowić podstawę dla automatycznej diagnostyki oraz monitorowania choroby Parkinsona.

Takie podejście niesie ze sobą wiele korzyści, które mogą rewolucjonizować sposób diagnozowania oraz monitorowania tej neurodegeneracyjnej choroby. Głos, będący wskaźnikiem stanu układu nerwowego i zdolności komunikacyjnych, dostarcza szeroką gamę informacji kluczowych dla procesu diagnozowania. Różnorodność parametrów akustycznych i fonacyjnych, które można zbadać, otwiera drzwi do kompleksowej oceny zmian zachodzących w organizmie pacjenta.

Wczesne objawy choroby Parkinsona często bywają trudne do wykrycia, szczególnie w standardowych badaniach klinicznych. Analiza głosu pozwala na szybką identyfikację subtelnych zmian, które pojawiają się we wczesnych fazach choroby. Ta wczesna detekcja umożliwia natychmiastową interwencję terapeutyczną, co może wpłynąć na spowolnienie progresji choroby i poprawę jakości życia pacjenta.

Analiza głosu jako narzędzie diagnostyczne wprowadza nowe perspektywy dla specjalistów zajmujących się chorobą Parkinsona. Logopedzi, foniatri i lekarze mogą wykorzystać obiektywne dane akustyczne do dokładnej oceny zmian w mowie i jakości głosu pacjenta. To również umożliwia sugerowanie odpowiednich interwencji, włącznie z dostosowaniem leczenia farmakologicznego.

Dla pacjentów oznacza to bardziej konkretne oceny ich stanu i dostosowane terapie, przyczyniające się do zwiększenia efektywności procesu leczenia. Badanie to jest szybkie, wygodne i bezpieczne, co może zachęcać pacjentów do systematycznego uczestnictwa w procesie diagnostycznym. Wprowadza to szczególną wartość dla pacjentów, którym trudno się przemieszczać.

Analiza głosu jako narzędzie diagnostyczne przy chorobie Parkinsona otwiera drzwi ku nowym, zaawansowanym metodologiom diagnozowania i leczenia. Może być wykorzystana do oceny stanu pacjenta oraz sugerowania odpowiednich interwencji, włączając zmiany w leczeniu farmakologicznym. W konsekwencji ma potencjał stania się szybkim, nieinwazyjnym wsparciem diagnostycznym i terapeutycznym, przyczyniając się do polepszenia opieki nad pacjentami dotknietymi chorobą Parkinsona.

2.2. Metody diagnozowania i monitorowania

Diagnostyką choroby Parkinsona zajmują się neurologi i geriatricy. Rozwój PD jest długotrwały, a w początkowych latach klinicznie niemal niewidoczny, co utrudnia wczesne rozpoznanie. Subtelne objawy często są uważane za skutek starzenia się lub błędnie diagnozowane jako inne zaburzenia neurologiczne. Kluczowym elementem w tym stadium jest dokładny wywiad, badanie fizyczne oraz identyfikacja objawów przez lekarza. Następnie diagnoza jest rozwijana poprzez badania laboratoryjne oraz obrazowe. Niestety, wyniki tych badań rzadko potwierdzają diagnozę od razu.

Początkowo pacjent zwykle konsultuje się z lekarzem pierwszego kontaktu, który powinien dokonać wstępnej diagnozy i skierować do neurologa. W tej fazie diagnozy przeprowadza się szczegółowy wywiad, uwzględniający rodzaj, nasilenie oraz okres występowania objawów, a także obecność chorób neurozwyrodnieniowych w rodzinie. Neurolog przeprowadza kompleksowe badanie neurologiczne, identyfikując symptomy takie jak sztywność mięśni, ograniczenia w ruchu (spowolnienie, trudności w poruszaniu się), drżenia spoczynkowe (np. w głowie, palcach rąk) oraz zaburzenia postawy i równowagi (zgarbienie, niestabilność, upadki). Kolejne badania są wykonywane w celu potwierdzenia lub wykluczenia diagnozy [10, 11].

a) Badania laboratoryjne

Obecnie brak specyficznych badań laboratoryjnych krwi, które potwierdzałyby diagnozę choroby Parkinsona. Niemniej jednak takie badania są użyteczne w wykluczaniu innych chorób o podobnym przebiegu. Wykonuje się podstawowe badania, takie jak morfologia krwi, elektrolity, poziom glukozy, TSH, próby wątrobowe, mocznik, kreatynina oraz poziom witaminy B12.

b) Badania obrazowe

Badania obrazowe głowy są przeprowadzane w celu wykluczenia innych chorób o podobnych objawach. Zalicza się do nich tomografię komputerową, ultrasonografię mózgu (USG) oraz rezonans magnetyczny głowy (MRI). Międzynarodowe kryteria rozpoznania choroby Parkinsona nie nakładają obowiązku wykonywania badań obrazowych w celu potwierdzenia diagnozy. Dostępne są również zaawansowane techniki obrazowania, takie jak PET (pozytonowa emisyjna tomografia) oraz SPECT (tomografia emisyjna pojedynczego fotonu), które pozwalają na obserwację metabolizmu w układzie pozapiramidowym. Skan DAT (skan transportera dopaminy) jest przykładem SPECT i może być sugerowany przez specjalistę. Mimo to ostateczna diagnoza opiera się na objawach oraz wynikach badania neurologicznego, a DAT nie jest wymagany.

c) Test z lewodopą

Test polega na podaniu pacjentowi preparatu z lewodopą. Jeśli następuje poprawa po zazyciu, istnieje wysokie prawdopodobieństwo, że pacjent rzeczywiście cierpi na chorobę Parkinsona. W przypadku braku poprawy, konieczne może być dalsze rozszerzenie diagnostyki.

d) Badania genetyczne

Choroba Parkinsona może występować rodzinie, co skłania do rozważenia diagnostyki genetycznej u pacjenta i jego krewnych. Obecnie zidentyfikowano 12 mutacji genów, które mogą wpływać na ryzyko zachorowania na PD, należy jednak pamiętać, że badania genetyczne są kosztowne.

e) Badania węchu

Większość osób z chorobą Parkinsona (90%) doświadcza zaburzeń węchu, manifestujących się hiposmią (osłabienie węchu), także we wczesnym stadium choroby. Jednak nie obserwuje się tych zaburzeń w przypadku zaniku wieloukładowego i następującego porażenia nadjądrowego.

f) Badania neuropsychologiczne i neuropsychiatryczne

Badania te służą identyfikacji zaburzeń poznawczych i emocjonalnych. Celem jest diagnoza łagodnych zaburzeń poznawczych, otępienia, a także zaburzeń psychotycznych, lękowych, kontroli impulsów i depresji.

Objawy przypominające chorobę Parkinsona są diagnozowane jako parkinsonizm i mogą być spowodowane różnymi zaburzeniami, takimi jak następujące porażenie nadjądrowe, zanik wieloukładowy, drżenie samoistne, demencja z ciałami Lewy'ego, choroby naczyniowe mózgu, otępienie, reumatyzm i inne [10]. Różnicowanie tych schorzeń jest kluczowe, ponieważ leczenie i podejście terapeutyczne są różne. Chociaż ostateczną diagnozę można ustalić tylko na podstawie badania mózgu po zgonie, wcześniej zdefiniowane kryteria diagnostyczne pozwalają na dokonanie diagnozy klinicznej. Według badania z 2021 roku opisanego w publikacji [12] diagnoza klinicznie potwierdzonej choroby Parkinsona może zabierać od kilku miesięcy do kilku lat, zależnie od indywidualnych czynników oraz reakcji na terapię lewodopą. Uznaje się, że aktualnie wykorzystywane kryteria pozwalają na diagnozę z dokładnością 90%, ale trwa to średnio 2,9 roku.

Pomocnym narzędziem w diagnostyce są szeroko stosowane skale oceny choroby Parkinsona. Stanowią istotne narzędzie w monitorowaniu stanu pacjentów oraz pomagają lekarzom i opiekunom ocenić stopień nasilenia objawów ruchowych, jak również wpływ choroby na codzienne funkcjonowanie pacjenta. Popularne skale, takie jak Skala Hoehn-Yahra, Skala UPDRS (Unified Parkinson's Disease Rating Scale) oraz Skala Schwab-England, umożliwiają stosunkowo obiektywną analizę symptomów i wsparcie w podejmowaniu decyzji terapeutycznych. Dzięki tym narzędziom możliwe jest dostosowanie leczenia do indywidualnych potrzeb pacjenta oraz śledzenie skuteczności terapii na przestrzeni czasu.

Obecnie proces diagnozy jest wyjątkowo złożony i wieloetapowy. W odpowiedzi na te wyzwania naukowcy koncentrują się na opracowaniu bardziej efektywnych narzędzi diagnostycznych. Rozwinięcie skuteczniejszych narzędzi diagnostycznych przyniesie korzyści nie tylko finansowe, ale także pozwoli na szybsze i trafniejsze udzielanie pomocy pacjentom. W nadchodzących latach, dążenie do wypracowania bardziej efektywnych metod diagnozowania choroby Parkinsona będzie kluczowym krokiem w zapewnieniu lepszej opieki zdrowotnej i poprawie jakości życia pacjentów.

2.3. Terapia osób chorych

Obecnie brak jest kuracji na chorobę Parkinsona, dlatego terapia skupia się na przywracaniu pacjentom zdolności funkcjonowania lub, w przypadkach zaawansowanych, na poprawie jakości życia. Zgodnie z aktualnym standardem medycznym, w terapii wykorzystuje się różnorodne metody, w tym leczenie farmakologiczne, głęboką stymulację mózgu oraz rehabilitację [3].

Leczenie farmakologiczne choroby Parkinsona opiera się na zwiększeniu poziomu dopaminy w mózgu, co wpływa na kontrolę objawów ruchowych i niezwiązanych z ruchem. Główną terapią jest lewodopa, która jest przetwarzana przez komórki nerwowe w dopaminę. Leczenie lewodopą często łączy się z karbidopą, która zmniejsza skutki uboczne i ilość potrzebnej lewodopy. Stosuje się też inne terapie farmakologiczne o różnych zasadach działania m.in. pobudzające produkcję dopaminy, zwiększające ilość dopaminy poprzez spowolnienie jej rozkładu, redukujące ruchy mimowolne czy zmniejszające drżenie i sztywność mięśni.

W przypadku pacjentów, u których leczenie farmakologiczne nie przynosi oczekiwanych efektów, może być rozważana Głęboka Stymulacja Mózgu (ang. *Deep Brain Stimulation-DBS*). W tym procederze chirurgicznym lekarz implantuje elektrody w określone obszary mózgu, łącząc je z małym urządzeniem elektrycznym umieszczonym w klatce piersiowej. Poprzez bezbolesne stymulowanie konkretnych obszarów mózgu kontrolujących ruch, DBS może pomóc w zmniejszeniu wielu objawów związanych z ruchem, takich jak drżenie, spowolnienie ruchu i sztywność.

Kluczową rolę w leczeniu odgrywa rehabilitacja neurologiczna, rozpoczynając się już od momentu postawienia diagnozy. Jej wsparcie jest nieocenione w łagodzeniu zaburzeń chodu, głosu, drżenia, sztywności oraz pogorszenia funkcji umysłowych. Wśród różnorodnych terapii, znajdują się między innymi:

- zbilansowana dieta: wspiera ogólne samopoczucie pacjenta,
- ćwiczenia fizyczne: wzmacniają mięśnie, poprawiają równowagę, elastyczność i koordynację,
- masaż terapeutyczny: pomaga w redukcji napięcia mięśniowego oraz przynosi ulgę w objawach,
- joga i tai-chi: wspomagają rozciąganie i elastyczność ciała, wpływając korzystanie na zdolność ruchową,
- rehabilitacja foniczna: pomaga w eliminowaniu trudności w mówieniu,
- psychoterapia: zapewnia wsparcie i umożliwia pacjentom pełne cieszenie się życiem pomimo choroby.

W ostatnim czasie na rynku pojawiło się też wiele aplikacji, które mają poprawić jakość życia osób z chorobą Parkinsona. Na przykład, *Parkinson's Central* zawierająca informacje dla pacjentów i opiekunów, obejmujące objawy, leczenie, wizyty lekarskie, zdrowy styl życia, badania i inne aspekty. Podobną aplikacją jest *Parkinson Symptom Tracker (PRO-PD App)*, zaprojektowana jako narzędzie do oceny i monitorowania nasilenia objawów choroby Parkinsona w czasie. Bazuje na wywiadzie w formie testu, który umożliwia ocenę objawów. Skala została opracowana w celu bycia wrażliwą na różne etapy choroby, charakteryzuje się dokładnością, szybkością, prostotą użycia i dostępnością.

Również w dziedzinie rehabilitacji ruchowej istnieje kilka aplikacji, takich jak *Lift Pulse*, służąca do rejestrowania danych dotyczących drżenia rąk. *Delay the Disease*, czyli program fitnessu rozwijany przez OhioHealth dla osób z chorobą Parkinsona, mający na celu poprawę funkcji fizycznych i opóźnienie postępu objawów choroby. W kontekście rehabilitacji poprzez gry cyfrowe warto wspomnieć o platformie *MindMotion® GO*. Oferuje ona różnorodne aktywności i śledzenie ruchu ciała dzięki technologii śledzenia markerów. Jedną z istotnych zalet programu to zdalne monitorowanie i dostosowanie przez terapeutę, zapewniające efektywną opiekę.

Obok aplikacji rehabilitacyjnych istnieje również *Parkinson's Cognitive Research* dedykowana osobom zainteresowanym uczestnictwem w badaniach naukowych dotyczących objawów poznawczych związanych z chorobą Parkinsona. Aplikacja umożliwia analizę aspektów takich jak skupiona uwaga, percepcja wzrokowa, rozpoznawanie, pamięć krótkotrwała, krótkotrwała pamięć wzrokowa, nazewnictwo, pamięć operacyjna, elastyczność poznawcza, planowanie, czas reakcji i prędkość przetwarzania. Jej głównym celem jest wspieranie badań naukowych poprzez dostarczanie narzędzi cyfrowych do oceny i terapii poznawczej. Mimo że stanowi cenny instrument dla społeczności naukowej oraz uniwersytetów na całym świecie, to jednak ma wyłącznie charakter badawczy i nie jest przeznaczona do diagnozowania ani leczenia choroby Parkinsona.

2.3.1. Rola i zastosowanie aplikacji głosowych w poprawie jakości mowy u osób z chorobą Parkinsona

Tematem przewodnim tej pracy magisterskiej jest rola głosu w kontekście choroby Parkinsona. W związku z tym przedstawione zostaną różnorodne aplikacje dostępne obecnie na rynku, które koncentrują się na poprawie jakości głosu w przypadku tej choroby.

Jedną z nich jest *Speak Up For Parkinson's*, przedstawiona przez Northwest Parkinson's Foundation. Aplikacja skupia się szczególnie na głośności głosu pacjenta. Oferuje dwa narzędzia do ćwiczeń: *Słowa i Zwroty*, gdzie użytkownik musi wypowiedzieć serię losowych stwierdzeń oraz *Czytanie i Konwersacja*, czyli obszar do swobodnych ćwiczeń o dłuższym czasie trwania. W obu narzędziach dostarczany jest miernik głośności oraz opinie dźwiękowe/wideo.

Inną propozycją jest aplikacja *Voice Trainer*, stworzona z myślą o osobach cierpiących na problemy z mową związane z chorobą Parkinsona, ale również jest przydatna dla profesjonalnych mówców oraz, przy wsparciu logopedy, dla osób z innymi zaburzeniami głosu lub mowy. W ramach aplikacji wyświetlane są wizualne informacje zwrotne dotyczące głośności i tonu mowy za pomocą jednej kropki na ekranie, co pozwala szybko zidentyfikować obszary wymagające poprawy. Aplikacja może być używana zarówno do ćwiczenia techniki, jak i do monitorowania toku rozmowy.

Kolejną godną uwagi aplikacją jest *Delayed Auditory Feedback (DAF)*. Przeznaczona jest dla osób z zaburzeniami mowy, które charakteryzują się szybkim tempem wypowiedzi. Aplikacja pomaga użytkownikom zwolnić tempo mówienia, co w efekcie czyni je bardziej zrozumiałymi dla innych. Działanie opóźnionej zwrotności dźwiękowej (DAF) polega na zmienionym odbiorze własnej mowy. Zakłócenie

normalnego cyklu zwrotnej informacji dźwiękowej skutkuje spowolnieniem tempa mówienia i poprawą klarowności wypowiedzi.

Aplikacja *Beats Medical Parkinson's* obejmuje rehabilitację objawów związanych z mową, chodem oraz drżeniem rąk. Została zaprojektowana tak, aby pomóc użytkownikom ćwiczyć głośne i wyraźne mówienie, co przekłada się na większą pewność siebie podczas komunikacji. Ważnym atutem aplikacji jest możliwość otrzymywania informacji zwrotnych w czasie rzeczywistym podczas ćwiczeń. Dzięki temu użytkownicy mogą śledzić swoje postępy i dostosowywać trening do swoich potrzeb.

Aplikacja LSVT LOUD skupia się na treningu osób z chorobą Parkinsona w celu osiągnięcia bardziej naturalnego poziomu głośności podczas mówienia w codziennych sytuacjach, takich jak komunikacja w domu, pracy czy w społeczności. W kontekście teorii terapeutycznej kluczowym aspektem tej metody jest pomoc pacjentom w *rekalibracji* ich percepji dźwięku, aby byli świadomi własnego głosu w kontekście interakcji z innymi ludźmi. Dzięki tym treningom, osoby z chorobą Parkinsona mogą lepiej kontrolować swoją głośność mówienia i efektywniej uczestniczyć w komunikacji.

Zupełnie inną aplikacją, zapewniającą ciągłe wsparcie ze strony specjalistów jest *Teleatherapy*. Oferuje ona terapię głosu dla osób z chorobą Parkinsona. Dzięki dostępowi zdalnemu pacjenci mogą oceniać swój stan oraz cele terapeutyczne, monitorować postępy i otrzymywać spersonalizowane wsparcie od logopedów. Aplikacja kieruje użytkowników przez program terapii mowy, a ćwiczenia są dostosowane do indywidualnych potrzeb i przypisywane przez logopedę. Ponadto terapeuci mają możliwość monitorowania ćwiczeń i udzielania opinii zwrotnej.

Aplikacja *Voiceitt* nie ma zastosowania w rehabilitacji, ale może znacznie poprawić komfort życia osób z chorobą Parkinsona. Pomaga tłumaczyć niewyraźne lub nietypowe dźwięki na zrozumiałą mowę. Każdy użytkownik trenuje oprogramowanie, podążając za prostymi komendami na ekranie. Aplikacja może być zintegrowana jako samodzielne ASR (*Automatic Speech Recognition*) dla osób o nietypowej mowie lub używany z innym systemem ASR jako rozszerzenie dostępności z dodatkowym interfejsem dla osób z ograniczoną kontrolą ruchową i zaburzeniami mowy. Planowane jest rozszerzenie aplikacji o napisów na spotkaniach po asystentów głosowych i sterowanie inteligentnym domem.

Powyższy przegląd ukazuje różnorodność dostępnych aplikacji, skierowanych do osób z chorobą Parkinsona. Te narzędzia cyfrowe stanowią cenną pomoc dla pacjentów, jednak istotne jest zrozumienie, że nie zastępują one profesjonalnej terapii mowy prowadzonej przez certyfikowanych logopedów. Ich rola polega na ułatwianiu ćwiczeń i treningów, ale nie mogą pełnić funkcji zastępczej dla terapii specjalistycznych. Warto zaznaczyć, że żadna z omówionych aplikacji nie ma zdolności diagnostycznych. Ich działanie opiera się na wsparciu w ćwiczeniach, terapii i monitorowaniu postępów, ale nie mają zdolności stawiania diagnoz.

Mimo tych ograniczeń, rosnące zainteresowanie dziedziną terapii mowy w kontekście choroby Parkinsona jest obiecujące. Wspomniane aplikacje są dowodem na rozwijający się obszar badań i innowacji w tej dziedzinie. W miarę postępu technologicznego istnieje szansa na powstanie nowych, bardziej zaawansowanych produktów, które mogą jeszcze skuteczniej wspomagać osoby z chorobą Parkinsona w poprawie jakości mowy i komunikacji.

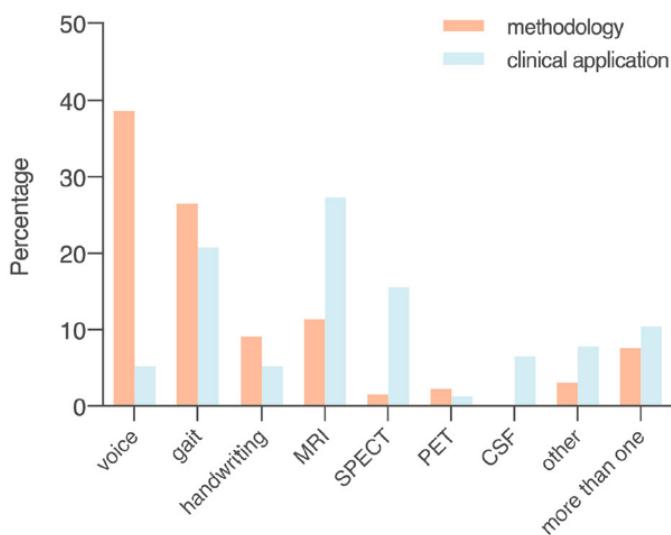
3. Analiza rozwiązań do automatycznej diagnostyki choroby Parkinsona

Diagnoza PD jest powszechnie oparta na obserwacjach lekarskich i ocenie objawów klinicznych, w tym charakterystycznego różnorodnych objawów ruchowych. Rosnąca liczba zachorowań i obniżenie wieku osób będących w grupie ryzyka skutkuje wzrostem zainteresowania dotyczącym narzędzi, które ułatwiałyby zarówno codzienne funkcjonowanie pacjentów, jak i pracę lekarzy. Tradycyjne metody diagnostyczne mogą być obarczone subiektywizmem, ponieważ opierają się między innymi na ocenie ruchów, które są czasami subtelne dla ludzkiego oka i dlatego trudne do sklasyfikowania, co może przyczynić się do błędnej diagnozy. Ponadto wczesne objawy niemotoryczne PD mogą być łagodne oraz spowodowane wieloma innymi schorzeniami. Dlatego też rozpoznanie tej choroby na wczesnym etapie stanowi wyzwanie.

Sztuczna inteligencja oraz nowoczesne technologie coraz częściej stają się integralną częścią systemu ochrony zdrowia. Wspierają lekarzy podczas diagnozy oraz wyboru sposobu leczenia pacjenta, a także pozwalają na monitorowanie choroby. Aby rozwiązać trudności i udoskonalić procedury diagnostowania oraz oceny PD, poszerzono stan techniki o metody uczenia maszynowego do klasyfikacji PD i osób zdrowych lub pacjentów z podobnymi objawami klinicznymi (np. zaburzeniami ruchu lub innymi zespołami parkinsonowskimi).

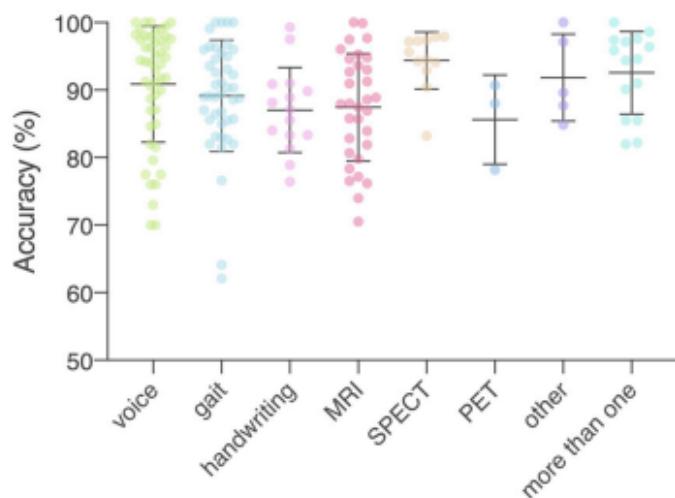
3.1. Rodzaj wykorzystywanych danych

Diagnozowanie choroby Parkinsona stanowi zadanie złożone ze względu na różnorodność objawów, które dotykają różnych aspektów funkcjonowania ciała i umysłu ludzkiego. W związku z tym, techniki uczenia maszynowego wykorzystywane w tym obszarze, także skupiają się na różnych rodzajach danych. Wśród źródeł informacji diagnostycznej znajdują się wyniki badań obrazowych (np. rezonans magnetyczny (MRI), tomografia komputerowa (SPECT)), które wydają się intuicyjne, biorąc pod uwagę zmiany w aktywności mózgu, które można zaobserwować. Niemniej jednak istnieją także mniej oczywiste metody diagnozy, które budzą duże zainteresowanie w środowisku naukowym, szczególnie w początkowym stadium choroby. Przykłady to analiza głosu, ocena charakterystyki chodu oraz badanie pisma odręcznego.



Rys. 3.1. Wykorzystanie rozwiązań w opracowaniach teoretycznych i zastosowaniach klinicznych w zależności od rodzaju danych (stan na 14 lutego 2020) [13]

Rys. 3.1 ilustruje zastosowanie wymienionych metod zarówno w teorii, jak i praktyce. Metody oparte na obrazowaniu medycznym wykazują wyraźną przewagę w zastosowaniach klinicznych w porównaniu do kontekstu teoretycznego. Niemniej jednak to pozostałe metody budzą znacznie większe zainteresowanie ze strony środowiska naukowego. Szczególnie w przypadku analizy głosu, gdzie rozbieżność między teorią a praktyką jest szczególnie znacząca. Przyczyny tego zjawiska zostaną dokładniej rozważone w dalszej części pracy. Co ciekawe, jak przedstawiono na Rys. 3.2, detekcja choroby na podstawie głosu daje bardzo wysokie wyniki, w większości analizowanych artykułów.



Rys. 3.2. Dokładność rozwiązań diagnostycznych w zależności od rodzaju danych (stan na 14 lutego 2020) [13]

Niniejsza praca dotyczy diagnostyki opartej na analizie głosu, dlatego temat ten zostanie bliżej rozważony. Założeniem dla takich systemów jest zadanie potencjalnemu pacjentowi zadania wokalnego, mogą to być:

- podtrzymywane samogłoski (ang. *sustained vowels*),
- zadanie diadochokinetyczne (DDK), mogące mierzyć zdolność do wydawania serii szybkich i naprzemiennych dźwięków (sylab),
- czytanie tekstu,
- wypowiedzenie pojedynczego zdania,
- monolog.

Dotychczas nie przeprowadzono badań porównawczych dotyczących wpływu wyboru zadania wokalnego na efektywność klasyfikacji w przypadku choroby Parkinsona (PD). Jedynie w artykule [14] przeprowadzono takie porównanie, przy użyciu cech prozodii (brzmieniowych właściwości mowy nakładających się na głosowy, sylabiczny i wyrazowy ciąg wypowiedzi). Nie pozwala to jednak na wyciągnięcie obiektywnych wniosków, ponieważ każde z zadań wokalnych może wymagać innego podejścia i metod klasyfikacji. Niemniej jednak, autorzy publikacji [15] przeprowadzili badanie na 200 pacjentach z PD, gdzie dokonano klasyfikacji deficytów mowy na pięć poziomów nasilenia. Oceniono typ (głos, artykulacja, płynność) oraz zakres upośledzenia dla każdego poziomu, korzystając z 2-minutowego fragmentu mowy. Wyniki ukazały, że głos stanowił najczęściej występujący i bardziej nasiłony deficyt we wczesnych stadiach choroby. Deficyty artykulacji i płynności pojawiały się później. Wykazano, że upośledzenie artykulacji korelowało z upośledzeniem głosu w fazie *cięzkiej*, a w fazie *głębokiej* dominującą cechą była upośledzona artykulacja.

W rezultacie, w kontekście wczesnej diagnostyki choroby, artykulacja i płynność mowy nie wymagają głębokiej uwagi. Koncentrację należy skupić przede wszystkim na cechach głosowych, co sprawia, że wybór podtrzymywanych samogłosek jako zadania wokalnego wydaje się najlepszym wyborem ze względu na ich stabilność w czasie oraz łatwość wypowiadania przez pacjenta. To podejście może posiadać potencjał uniwersalności dla różnych języków, co oznacza, że analiza może być stosowana niezależnie od języka ojczystego pacjenta. W efekcie pozwala to na bardziej efektywne i znormalizowane diagnozowanie choroby Parkinsona.

Najczęściej występującymi w opracowaniach samogłoskami są /a/, /e/, /i/ oraz /u/. Aktualnie brakuje albo nie udało się jeszcze ustalić, która z tych samogłosek niesie najcenniejsze informacje z punktu widzenia diagnostycznego. W związku z tym, w ramach badawczej części niniejszej pracy przeprowadzone zostanie takie porównanie dla wybranych metod klasyfikacji.

3.2. Metody klasyfikacji

W obszarze automatycznej diagnostyki choroby Parkinsona (PD) opartej na analizie głosu, istnieje wiele różnych metod klasyfikacji. W publikacji [13] przeanalizowano 55 artykułów dotyczących tego

zagadnienia, które zostały opublikowane przed lutym 2020 roku. Analiza ujawniła, że średnia dokładność osiągnięta w tych badaniach wyniosła 90,9%, z zakresem wyników od 70,0% [16, 17] do 100,0% [18, 19, 20, 21]. Wyniki były zależne od różnorodnych zbiorów danych oraz różnych podejść do analizy głosu.

Istnieje wiele czynników, które wpływają na wybór konkretnej metody klasyfikacji w zależności od kontekstu diagnostycznego. Przykładowo, różne metody mogą być skuteczniejsze w analizie mowy spontanicznej w porównaniu do mowy kontrolowanej. W niniejszej pracy skupiono się na metodach związanych z analizą podtrzymywanych samogłosek, co stanowi główny obszar badawczy.

Tradycyjna inżynieria cech to jedno z najpowszechniejszych podejść w automatycznej diagnostyce choroby Parkinsona na podstawie analizy głosu. To podejście polega na wydobyciu charakterystycznych cech z sygnału mowy, takich jak formanty, nieregularność odstępów czasowych (*jitter*), współczynniki cepstralne MFCC, stosunek poziomu hałasu do poziomu harmonicznych składników (NHR), stosunek poziomu harmonicznych składników do poziomu szumu (HNR), wskaźnik migotania *Shimmer*, częstotliwość podstawowa (F0) i wiele innych. Choć to podejście jest popularne, istnieją również bardziej zaawansowane metody, które mogą dostarczyć dokładniejszych wyników. Niemniej jednak tradycyjna inżynieria cech ma swoje zalety, takie jak możliwość interpretacji wyników przy użyciu powszechnie znanych miar akustycznych.

Inną perspektywą jest wykorzystanie spektrogramów jako podstawy dla procesu klasyfikacji. Proces ten polega na przekształceniu dźwięków mowy na formę wizualną na przykład w postaci spektrogramów, które obrazują zmiany w czasie i częstotliwości. Następnie, sieci konwolucyjne (CNN), które są zaprojektowane do pracy z obrazami, mogą analizować struktury i wzorce w spektrogramach. To podejście może umożliwić dokładniejszą klasyfikację, szczególnie dla cech trudnych do wykrycia za pomocą innych metod.

W publikacji [22] przeprowadzono porównanie tradycyjnego podejścia opartego na inżynierii cech z nowoczesnym wykorzystaniem głębokich sieci neuronowych. Wyniki sugerują, że połączenie tych podejść może przynieść skuteczniejsze wyniki. Chociaż tradycyjna inżynieria cech osiągała lepsze wyniki, głębokie sieci, takie jak te z rodzin VGG i ResNet, mają potencjał do osiągnięcia jeszcze lepszych rezultatów.

Publikacji na temat wykorzystania spektrogramów oraz CNN w klasyfikacji PD na podstawie mowy jest znacznie mniej niż tych dotyczących tradycyjnej inżynierii cech. Spośród 55 artykułów analizowanych w [13] tylko jeden dotyczył takiego podejścia. Jest to publikacja z 2019 roku opisana w [23]. Badanie wprowadziło głębokie sieci neuronowe typu CNN do klasyfikacji choroby Parkinsona na podstawie cech głosowych. Stworzono dwie struktury sieciowe, różniące się w sposobie przetwarzania cech. Pierwsza struktura łączyła różne zestawy cech przed ich analizą w 9-warstwowej sieci CNN. Druga struktura wykorzystywała równolegle wejścia, by jednocześnie ekstrahować głębokie cechy z różnych zestawów cech. Wyniki wykazały, że cechy TQWT (ang. *Time–Frequency Warping Transform*, rodzaj cech używanych do analizy dźwięków, uwzględniających ich zmienne tempo w czasie i częstotliwości) były najbardziej skuteczne w klasyfikacji. Połączenie ich z innymi cechami poprawiło wyniki. Model

drugiej struktury osiągał lepsze rezultaty niż SVM, zwłaszcza z łączением 3 zestawów cech. Zaproponowane podejście CNN osiągnęło skuteczność 86,9%, przewyższając inne metody. Są to pierwsze obiecujące rezultaty podejścia opartego na głębokich sieciach neuronowych.

Od 2020 roku ukazało się kilka dodatkowych publikacji, które pogłębiły ten temat. Na przykład w 2022 roku autorzy artykułu [24] wykorzystali spektrogramy oraz metodę wielokrotnego dopasowywania modelu. Model był wstępnie trenowany na zbiorze ImageNet, następnie adaptowany do pośredniego zbioru, a na koniec dostosowywany do danych PC-GITA (baza danych zawierająca nagrania samogłosek od osób posługujących się kolumbijskiej odmianie języka hiszpańskiego). Mimo niewielkich różnic pomiędzy różnymi samogłoskami najlepszą skuteczność osiągnięto przy uwzględnieniu samogłoski /a/. Uzyskano 99% dokładności, 86,2% czułości, 93,3% swoistości. Co ciekawe, wykazano, że skuteczność podejścia nie zależy od płci. To pokazuje, że metoda ma potencjał do zastosowania w praktyce klinicznej do przesiewowego badania, diagnozowania i monitorowania choroby Parkinsona.

Ciekawą analizę, również opartą na spektrogramach, opisano w publikacji [25]. Zaproponowano trzy podejścia - pierwsze, wykorzystujące transfer learning; drugie, wykorzystujące głębokie cechy wyodrębnione ze spektrogramów mowy za pomocą klasyfikatorów uczenia maszynowego; oraz trzecie, oceniające proste cechy akustyczne nagrań również przy użyciu klasyfikatorów uczenia maszynowego. Wyniki wskazują, że druga propozycja wykazuje obiecujące rezultaty. Zaobserwowano najwyższą dokładność na poziomie 99,7% dla samogłoski /o/ oraz odczytywanego tekstu przy użyciu perceptronu wielowarstwowego. Natomiast przy wykorzystaniu głębokich cech samogłoski /i/ uzyskano dokładność wynoszącą 99,1% przy użyciu lasu losowego. Z badania można wywnioskować, że metoda bazująca na głębokich cechach wykazuje lepsze wyniki w porównaniu do prostych cech akustycznych i podejść opartych na transfer learning.

Oprócz konwolucyjnych sieci neuronowych, w analizie spektrogramów wykorzystuje się również ELM (ang. *Extreme Learning Machines*). ELM to technika uczenia się maszynowego, w której warstwa wejściowa modelu jest inicjalizowana losowo, a wagę warstwy wyjściowej są wyznaczane poprzez rozwiązywanie jednokrotnego równania liniowego. ELM ma zalety efektywności obliczeniowej oraz zdolności do radzenia sobie z różnymi typami danych, w tym danymi wizualnymi i dźwiękowymi. W badaniu [26] porównano ELM oraz CNN. Otrzymane wyniki mieściły się między 81,74% a 83,91% dokładności. Analiza pokazała, że większa liczba próbek wpływa na lepsze wyniki, a sieć AlexNet miała najlepszą równowagę między rozproszeniem a wydajnością. W innym badaniu [27] przeanalizowano różne wersje ELM w celu klasyfikacji pacjentów z chorobą Parkinsona. Wyniki wskazują, że wielowarstwowe sieci ELM wykazują lepszą wydajność niż jednowarstwowe. Osiągnięto dokładność, oscylującą w okolicach 80%.

Sieci CNN używają metod konwolucyjnych do analizy relacji między funkcjami. W przeciwnieństwie do sieci rekurencyjnych sieci te są zazwyczaj stosowane do klasyfikacji obrazów i nie uwzględniają relacji sekwencyjnych. Dlatego w publikacji [28] zastosowano podejście, gdzie wyodrębnione cechy z sieci CNN są przekazywane do warstwy LSTM, aby nauczyć się informacji czasowych w dźwiękach, rozpoznawać sekwencyjne informacje i analizować stan choroby Parkinsona. Wykorzystano połączenie

modeli ResNet LSTM. Modele ResNet służą do wyciągania cech z obrazów mel-spektrogramu sygnałów mowy, a sieć LSTM jest wykorzystywana do rozpoznawania informacji sekwencyjnych z uzyskanych cech. Najwyższą wydajność klasyfikacji osiągnięto na poziomie 98,61%. Porównanie zaproponowanego modelu z aktualnym stanem wiedzy pokazuje jego wysoką wydajność w detekcji choroby Parkinsona.

Problemem, który często uniemożliwia uzyskanie zadowalających rezultatów, jest ograniczony rozmiar zbioru danych. Większość publicznie dostępnych baz danych składa się z około 50 nagrań, co jest niewystarczające do uzyskania wiarygodnych wyników i realnego zastosowania w medycynie klinicznej. Specyfika choroby utrudnia też samodzielne rozszerzenie bazy o nagrania pacjentów z chorobą Parkinsona. To wyzwanie stawiane jest przed każdym, kto próbuje stworzyć automatyczne narzędzie do diagnostowania choroby Parkinsona na podstawie analizy głosu. W artykule [29] przedstawiono wykorzystanie *Spectrogram-Deep Convolutional Generative Adversarial Network* (S-DCGAN) do augmentacji próbek głosowych, co może częściowo rozwiązać problem i przyczynić się do większej różnorodności w zbiorach danych. Na zestawie danych Sakar, hybrydowy model S-DCGAN-ResNet50 osiągnął najwyższą dokładność rozpoznawania wzorca głosowego wynoszącą 91,25% oraz swoistość na poziomie 92,5%, co pozwala na precyzyjniesze różnicowanie między pacjentami z PD a zdowymi osobami w porównaniu z modelem DCGAN-ResNet50.

Wszystkie te badania prowadzą do wniosku, że wykorzystanie uczenia maszynowego oraz głębokich sieci neuronowych może istotnie poprawić precyzję i skuteczność diagnostyki choroby Parkinsona na podstawie analizy cech głosowych. Te nowoczesne techniki otwierają nowe horyzonty w kwestii doskonalenia opieki nad pacjentami oraz procesów diagnostycznych w dziedzinie medycyny. Jednakże warto zaznaczyć, że wyniki te są osiągane na różnorodnych zbiorach danych, co utrudnia bezpośrednie porównania między poszczególnymi badaniami.

Najlepsze wyniki skuteczności klasyfikacji osiągnięto przy użyciu różnych podejść i modeli. Przykładem jest metoda opisana przez Gómeza et al. [25], która osiągnęła dokładność 99,7% dla jednej samogłoski. Z drugiej strony, istnieją też przypadki, jak w badaniu Guatelliego et al. [26], gdzie osiągnięto wyniki między 81,74% a 83,91% dokładności klasyfikacji, co może być związane z ograniczonym zbiorem danych. Konieczne jest dalsze prowadzenie prac w tym kierunku, najlepiej na jak najbardziej rozbudowanych bazach danych.

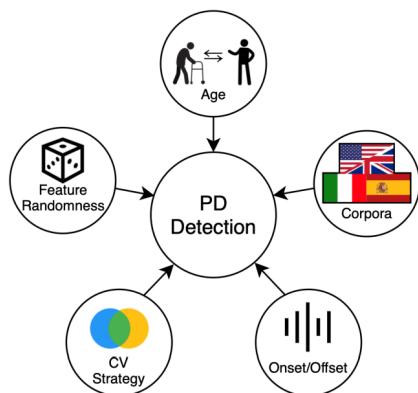
Wartością dodaną tych badań jest możliwość zastosowania analizy głosowej w procesie diagnostowania choroby Parkinsona, co może umożliwić wczesne wykrycie objawów i rozpoczęcie leczenia. Wyniki te sugerują, że głębokie sieci neuronowe mogą pomóc w identyfikowaniu subtelnych cech głosowych, które są trudne do wykrycia przez konwencjonalne metody diagnostyczne.

o tym [30] o tym [31] o tym [32] o tym [33]- dopisać

3.3. Wyzwania związane z systemami automatycznej diagnostyki

Zainteresowanie systemami do automatycznej diagnostyki choroby Parkinsona na podstawie głosu jest ogromne i wiąże się z nim duże nadzieje. Istnieje jednak duża dysproporcja pomiędzy pracą badawczą a ich wykorzystaniem w rzeczywistym środowisku (Rys. 3.1). Przyczyn takiego stanu rzeczy jest wiele, a większość z nich związana jest z brakiem usystematyzowanego podejścia do problemu, co utrudnia porównanie rozwiązań, a tym samym rzetelny postęp.

Ostatnie badania wykazały, że możemy wytrenować dokładne modele do wykrywania oznak PD z nagrań audio. Jednakże istnieją rozbieżności, które są częściowo powodowane przez różnice w wykorzystywanych korpusach lub metodologii. Dlatego autorzy publikacji [34] przeprowadzili analizę, wpływu niektórych czynników na wyniki klasyfikacji. Głównym celem była ich identyfikacja oraz stworzenie zasad, które w przyszłości pozwolą usystematyzować stan wiedzy w tej dziedzinie. W badaniach skupiono się na przedłużonych samogłoskach (ang. *sustained vowels*), ponieważ są one najlepszym i najpopularniejszym zadaniem wokalnym w takich systemach. Przeprowadzone eksperymenty wykazały, że nieuwzględnione zmienne w metodologii, projekcie eksperymentalnym i przygotowaniu danych prowadzą do zbyt optymistycznych wyników w badaniach nad automatyczną detekcją PD. Czynniki, które zidentyfikowano jako przyczyniające się do zbyt optymistycznych wyników klasyfikacji, przedstawiono na Rys. 3.3 oraz omówiono poniżej.



Rys. 3.3. Czynniki powodujące zbyt optymistyczne wyniki detekcji PD na podstawie głosu [34]

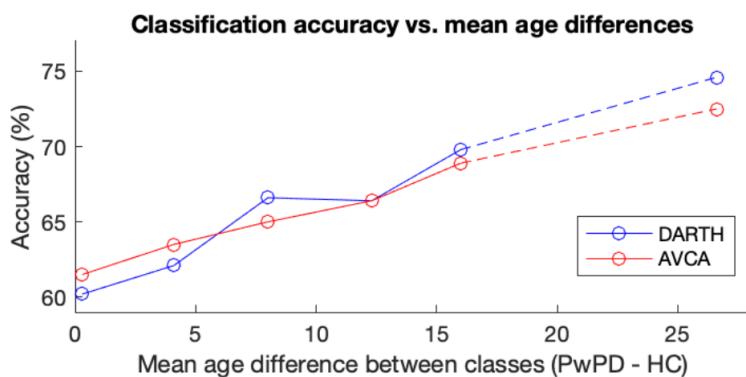
- Pominięcie aspektu tożsamości mówcy przy konstruowaniu zbiorów treningowych i testowych

W przypadku, gdy w zbiorze danych znajduje się kilka nagrań od tego samego mówcy, można postąpić na dwa sposoby. Pierwszy z nich to podział według podmiotów (ang. *subject-wise split*) polegający na tym, że nagrania od tej samej osoby znajdują się albo w zbiorze treningowym, albo testowym, natomiast nigdy w obu na raz. Drugie podejście to podział według rekordów (ang. *record-wise split*), gdzie nagrania są losowo dzielone do zbiorów lub intencjonalnie używa się nagrań od tej samej osoby zarówno w zbiorze testowym, jak i treningowym. Okazuje się, że podejście typu *record-wise* prowadzi do wyższej dokładności niż *subject-wise split*, jeśli pozostałe

założenia pozostają identyczne. Prawdopodobnie wynika to z faktu, że klasyfikator nastawia się na wykrywanie unikalnych informacji indywidualnych, reprezentowanych przez współczynniki takie jak MFCC, a nie rzeczywiste biomarkery lub wzorce PD. Dlatego też rekomendowana jest technika *subject-wise split*, aby uniknąć zbyt optymistycznych wyników.

b) Niezbalansowanie klas pod względem wieku

W literaturze można znaleźć prace wykorzystujące zbiory danych, w których średni wiek mówców w klasie osób chorych na PD różni się od średniego wieku w klasie osób zdrowych o ponad 5 lat. Autorzy zapewniają o wysokiej skuteczności swoich rozwiązań, jednak pomijają informacje o ryzyku, że klasyfikator uczy się wykrywać cechy powiązane z wiekiem, zamiast rzeczywistych wzorców PD. Wyniki eksperymentów w publikacji [34] pokazują, że wraz ze wzrostem różnicy między średnim wiekiem uczestników z PD i HC, dokładność klasyfikacji konsekwentnie rośnie (Rys. 3.4). Na tej podstawie można stwierdzić, że związany z wiekiem wpływ na głos mówców może zaburzać wyniki otrzymywane przez klasyfikator. Dlatego też zaleca się zbalansowanie używanych zbiorów danych, tak aby średnia różnica wieku między tymi dwoma klasami była jak najmniejsza.



Rys. 3.4. Wykres przedstawiający zależność różnicy wieku między klasami a dokładnością klasyfikacji [34]

c) wpływ losowości cech na dokładność klasyfikacji

W publikacji [34] przeprowadzono badania analizujące wpływ losowości cech na dokładność klasyfikacji. Zamieniono cechy obliczone za pomocą DARTH-VAT na losowe liczby, zachowując etykiety i podziały. Wyniki wskazały, że nawet losowe cechy mogą prowadzić do wysokich wyników klasyfikacji (ponad 72%). Efekt ten jest bardziej widoczny w mniejszych korpusach, gdzie różnica między liczbą nagrań a wymiarowością cech ma większy wpływ na potencjalną korelację przypadkową. Badanie pokazuje, że nadmierna liczba cech w stosunku do liczby obserwacji może prowadzić do fałszywie wysokich wyników klasyfikacji nawet przy użyciu losowych cech. Im większa różnica między liczbą plików a wymiarem wektora cech, tym większe szanse na znalezienie cechy, która losowo koreluje z etykietami klas. To sugeruje, że osiągnięcia klasyfikacyjne

powinny być analizowane w kontekście proporcji cech do próbki, aby uniknąć nadmiernie optymistycznych interpretacji wyników klasyfikacji w zastosowaniach medycznych.

d) ograniczenie losowego nadmiernego dopasowania poprzez uwzględnienie zbioru walidacyjnego

Dla mniejszych zbiorów danych, praktyką jest często używanie tylko zbiorów treningowych i testowych podczas krzyżowej walidacji. Jest to podejście, które może prowadzić do wyników zbyt optymistycznych, ponieważ wszystkie wyniki testowe są brane pod uwagę przy wyborze optymalnej konfiguracji modelu. Inną strategią jest wykorzystanie danych treningowych do oceny wytrenowanych modeli i późniejsze przetestowanie najlepszego modelu na zbiorze testowym. Niemniej jednak to podejście może być niepraktyczne, ponieważ może prowadzić do wyników idealnych (dokładność 100%) na zbiorze treningowym, co jest niepożądane. Aby uniknąć tych problemów, proponuje się wykorzystanie dodatkowego zbioru walidacyjnego [34]. Wybierając model na podstawie wyników walidacyjnych, a następnie testując go na zbiorze testowym, można uniknąć ryzyka nadmiernego dopasowania. Dla mniejszych zbiorów danych ta strategia może ograniczać dostępną liczbę danych treningowych, co wpływa na wydajność klasyfikacji.

e) wpływ początku i końca nagrani samogłosek na wyniki klasyfikacji

Główną różnicą między korpusami wykorzystywanyymi do klasyfikacji choroby Parkinsona jest obecność fragmentów nagrania oznaczonych jako *onset* i *offset*. Niektóre nagrania zawierają te segmenty, podczas gdy inne zostały ich pozbawione, aby zapewnić stabilniejszą fonację, co jest korzystne dla pewnych cech i algorytmów analizy. W celu oceny znaczenia informacji zawartych w obszarach *onset* i *offset* dla klasyfikacji przeprowadzono eksperymenty porównawcze, wykorzystując nagrania zarówno z fragmentami przyciętymi, jak i nieprzyciętymi [34]. Wyniki tych eksperymentów ukazały, że wyeliminowanie fragmentów początkowych i końcowych wpłynęło negatywnie na dokładność klasyfikacji. To wskazuje na to, że obszary te zawierają istotne informacje artykulacyjne, które mają znaczenie dla procesu klasyfikacji.

f) eksperymenty międzykorpusowe a zdolności generalizacyjne

Większość badań dotyczących diagnozowania choroby Parkinsona na podstawie głosu opiera się na jednym lub ewentualnie kilku (wykorzystywanych niezależnie) korpusach mowy. W tym kontekście często pomija się badanie zdolności klasyfikatorów do ogólnego zastosowania. W artykule [34] przeprowadzono międzykorpusowe eksperymenty na bazach danych w językach włoskim i hiszpańskim w celu przetestowania zdolności ogólnych modeli. Skuteczność tych modeli różniła się w zależności od języka zbioru testowego. Może to wynikać z odmiennej różnorodności nagrani, co ma wpływ na stabilność modelu. Drugim możliwym wyjaśnieniem jest to, że głos osób z chorobą Parkinsona może być w różnym stopniu obarczony objawami choroby w zależności od języka ojczystego lub stopnia zaawansowania choroby. Innymi słowy, w zależności od użytego zbioru danych objawy mogą być napisane w różny sposób i konieczne jest wzięcie tego pod uwagę tak by zdolności generalizacyjne modelu były jak najwyższe.

Identyfikacja i świadomość wpływu powyższych czynników pozwala na dostosowaniu przeprowadzanych eksperymentów tak, aby uniknąć wyników zbyt optymistycznych. Usystematyzowanie podejścia do analizy głosu pod kątem diagnostyki choroby Parkinsona przyczyni się do możliwości obiektywnego porównania istniejących i nowych rozwiązań. Tym samym przyspieszy to postęp w tej dziedzinie i uzyskanie optymalnego rozwiązania, które mogłoby zostać wykorzystane w rzeczywistym środowisku.

Nie są to jednak wszystkie czynniki, które zaburzają obiektywność wyników. Konieczna jest dyskusja na temat nowych kompleksowych linii bazowych dla prowadzenia eksperymentów w automatycznym wykrywaniu PD na podstawie fonacji, a także innych ogólnych zastosowań przetwarzania mowy.

Prace nad automatyczną detekcją Parkinsona na podstawie głosu trwają już od dłuższego czasu. Jednak wciąż brakuje systemu, który mógłby zostać uznany za wystarczająco niezawodne narzędzie diagnostyczne. Wśród problemów, które ograniczają rzeczywiste wykorzystanie takich systemów wyróżnia się:

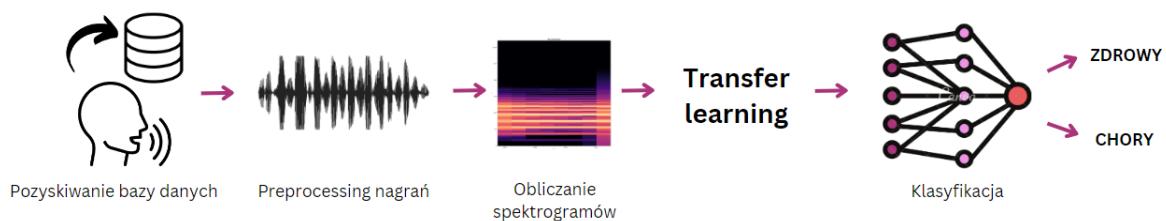
- zróżnicowanie wzorców mowy: osoby z chorobą Parkinsona mogą różnić się w sposób, w jaki zmiany w mowie wpływają na ich głos. To zróżnicowanie utrudnia stworzenie uniwersalnego modelu, który działałby skutecznie dla wszystkich pacjentów.
- wpływ zmiennych czynników: wpływ na mowę mogą mieć różne czynniki, takie jak zmęczenie, stres czy otoczenie akustyczne. Te zmienne mogą wprowadzać zakłócenia w analizie mowy i utrudniać jednoznaczną diagnozę.
- potrzeba dużej bazy danych: aby stworzyć dokładny system detekcji, konieczne jest posiadanie dużej bazy głosów osób z i bez choroby Parkinsona. Uzyskanie takiej bazy danych, która odzwierciedla różnorodność pacjentów i warunki środowiskowe, może być wyzwaniem. Większość publikacji opiera się na bazach danych zawierających około 50 nagrani, co nie jest wystarczająco reprezentatywną próbą.
- wcześnie wykrycie i subtelne objawy: wcześnie stadia choroby Parkinsona często manifestują się subtelnie, a różnice w mowie mogą być trudne do zauważenia. To może prowadzić do błędnych diagnoz lub niskiej skuteczności systemu.
- weryfikacja i walidacja: Aby narzędzie diagnostyczne oparte na mowie było skuteczne, musi być poddane rygorystycznym testom w rzeczywistych warunkach klinicznych. Weryfikacja i walidacja takiego systemu to skomplikowany proces.
- ograniczenia technologiczne: pomimo postępów w technologii analizy mowy, istnieją nadal ograniczenia w dokładności i precyzji takich systemów. Może to prowadzić do wyników fałszywie pozytywnych lub negatywnych.
- aspekty etyczne i prywatność: wykorzystywanie danych głosowych do diagnozowania chorób podnosi kwestie związane z prywatnością i etyką. Konieczne jest zagwarantowanie odpowiednich zabezpieczeń danych i zgody pacjentów na wykorzystanie ich informacji w celach medycznych.

Mimo tych wyzwań, prace nad wykorzystaniem analizy mowy do diagnozowania choroby Parkinsona są obiecujące i mogą przyczynić się do poprawy jakości życia pacjentów oraz usprawnienia procesu diagozy i leczenia. Jednakże przed stworzeniem skutecznego narzędzia diagnostycznego opartego na głosie jest jeszcze wiele pracy do wykonania.

W niniejszej pracy podjęta zostanie próba implementacji takiego rozwiązania. Uwzględnione zostaną wszystkie z rekomendacji przedstawionych w artykule [34].

4. Materiał i metoda badawcza

W ramach pracy przeprowadzone zostały badania dotyczące klasyfikacji choroby Parkinsona. Głównym celem było opracowanie efektywnego modelu klasyfikacji binarnej, który może rozróżniać osoby z PD od osób zdrowych na podstawie analizy sygnałów mowy. Przyjęte podejście opiera się na wykorzystaniu metod przetwarzania sygnałów mowy oraz uczenia maszynowego. Najpierw zebrano dane, w tym nagrania głosowe osób z chorobą Parkinsona (PD) oraz osób zdrowych (HC, ang. *Healthy Controls*). Następnie przeprowadzono analizę przy użyciu różnych ustawień spektrogramów i melspektrogramów. Obejmuje to zmienne parametry takie jak rozmiar okna i przesunięcie okna. Dodatkowo przeprowadzono badania dotyczące różnych architektur konwolucyjnych sieci neuronowych. Celem jest zbadanie, które architektury sieci i ustawienia spektrogramów dają najlepsze wyniki w klasyfikacji choroby Parkinsona dla poszczególnych samogłosek. Przeprowadzona analiza porównawcza pozwala na lepsze zrozumienie wpływu tych czynników na skuteczność klasyfikacji. W rezultacie możliwe będzie ustalenie optymalnych ustawień i architektur dla klasyfikacji choroby Parkinsona na podstawie analizy sygnałów mowy. Ponadto zbadano wpływ rozszerzenia zbioru danych o dodatkowe nagrania pochodzące od tych samych osób. Zostało to poglądowo przedstawione na Rys. 4.1.



Rys. 4.1. Schemat przyjętej metody badawczej

4.1. Materiał badawczy

Materiałem badawczym w niniejszej pracy magisterskiej są nagrania głosowe podtrzymywanych samogłosek (ang. *sustained vowels*) /a/, /e/, /i/, /o/ oraz /u/. Baza danych obejmuje nagrania osób zdrowych oraz cierpiących na chorobę Parkinsona. Podstawą skuteczności metod uczenia maszynowego jest odpowiednie przygotowanie danych, zarówno pod względem jakości, jak i ilości. To kluczowy element, który wpływa na wydajność i zdolność modelu do radzenia sobie w rzeczywistym środowisku. Dane użyte

do trenowania modelu muszą być dokładne, spójne i reprezentatywne dla rzeczywistego środowiska, w którym będzie działał model. Jest to podobne do procesu nauki lekarza – im więcej przypadków lekarz widzi i diagnozuje, tym lepiej radzi sobie z różnymi przypadkami w praktyce. W uczeniu maszynowym większa ilość danych pozwala na uchwycenie różnic międzyosobniczych i niuansów w danych, co przekłada się na większą stabilność i skuteczność modelu. Model musi być w stanie radzić sobie z nowymi danymi, które nie były używane podczas treningu.

W badaniu opublikowanym w artykule [34] podkreślono, że często zespoły badawcze zajmujące się rozwiązaniami do automatycznej diagnozy choroby Parkinsona na podstawie mowy polegają na jednej bazie danych. Takie podejście może prowadzić do wyników, które są zbyt optymistyczne. Nie uwzględniają one bowiem na przykład różnic związanych z pochodzeniem badanych oraz warunków i sposobu przeprowadzania nagrań.

Biorąc pod uwagę powyższe, podjęto decyzję o połączeniu trzech różnych baz danych w językach polskim, włoskim oraz hiszpańskim. Ta inicjatywa badawcza ma na celu zbadanie potencjalnej uniwersalności narzędzia służącego do automatycznej diagnozy choroby Parkinsona, tak by nie było ograniczone tylko do jednej grupy narodowościowej czy geograficznej. Pozwoliłoby to na rozwinięcie bardziej ogólnego i zastosowanego narzędzia diagnostycznego, które może być używane na skalę międzynarodową, przyczyniając się do poprawy jakości opieki zdrowotnej i diagnozy tej choroby.

Warto też zauważyć, że żadna z dostępnych publicznie baz danych nie zawiera wystarczającej ilości nagrań, by mówić o stabilnym rozwiązaniu, które mogłoby zostać użyte na szeroką skalę. Łączenie zestawów danych może znacząco poprawić zdolności generalizacyjne modelu.

4.1.1. Baza danych w języku polskim

Materiał badawczy pochodzący od osób cierpiących na chorobę Parkinsona (PD) i posługujących się językiem polskim, został zgromadzony podczas realizacji rozprawy doktorskiej we współpracy z Krakowskim Szpitalem Specjalistycznym im. Jana Pawła II [35]. Zawarta w bazie danych kolekcja obejmuje nagrania samogłosek /a/, /e/, /i/, /o/ oraz /u/ z wydłużoną intonacją, pozyskane od 27 pacjentów. Dla każdego z pacjentów dostępne jest jedno nagranie każdej z wymienionych samogłosek. W ramach przeprowadzonych badań pomiary zostały wykonane przed spożyciem leków oraz w określonych odstępach czasowych po podaniu lewodopy. Przed każdym pomiarem lekarz neurolog przeprowadzał badanie pacjenta i oceniał jego stan, wykorzystując skalę UPDRS. W niniejszej pracy magisterskiej wykorzystano jedynie nagrania głosowe zarejestrowane przed podaniem leku.

Nagrania osób zdrowych zostały zebrane w ramach tej pracy magisterskiej przy wykorzystaniu aplikacji *Easy Voice Recorder*. Ustalono protokół nagrywania, wykluczając osoby poniżej 50 roku życia, palące oraz ze zdiagnozowaną lub podejrzaną chorobą wpływającą na aparat mowy, lub korę mózgową (np. choroba Parkinsona, epilepsja, padaczka). Pozyskiwano nagrania jedynie od osób, dla których językiem ojczystym jest polski.

Każda osoba kwalifikująca się do badania otrzymała zadanie trzykrotnego wypowiedzenia samogłosek /a/, /e/, /i/, /o/ oraz /u/ w odstępach czasowych, utrzymując dźwięk przez co najmniej 2 sekundy.

Następnie nagrania zostały dokładnie przeanalizowane. Usunięto nagrania zbyt krótkie oraz te, które nie spełniały kryteriów dotyczących jakości. Otrzymana baza danych nadal zawierała nagrania od 27 osób chorych oraz od 26 osób zdrowych, zmieniła się jedynie liczbeność nagrań dla poszczególnych samogłosek. W tabeli 4.1 przedstawiono informacje dotyczące bazy danych.

Tabela 4.1. Charakterystyka polskiej bazy danych

Kategoria	Osoby zdrowe (HC)	Osoby chore (PD)	Razem
Liczba osób	26	27	53
Liczba nagrań	67	27	94
Średnia wieku	$60,88 \pm 7,98$	$64,49 \pm 8,49$	$62,68 \pm 8,43$
Liczba kobiet	18	13	31
Liczba mężczyzn	8	14	22

4.1.2. Baza danych w języku włoskim

Włoska baza danych to *Italian Parkinson's Voice And Speech*, która dostępna jest na platformie *IEEEDataPort* [36]. Zbiór zawiera wiele różnych podzbiorów, ale wykorzystano jedynie podtrzymywane samogłoski /a/, /e/, /i/, /o/ oraz /u/. Ocena stopnia zaawansowania choroby została przeprowadzona przy pomocy skali UPDRS. Wszystkie nagrania pochodzą od osób, dla których natywnym językiem jest włoski. Początkowo baza danych zawiera nagrania od 50 osób, jednak po wstępny oczyszczeniu wykorzystano jedynie 45. Charakterystykę zbioru przedstawiono w tabeli 4.2

Tabela 4.2. Charakterystyka włoskiej bazy danych

Kategoria	Osoby zdrowe (HC)	Osoby chore (PD)	Razem
Liczba osób	19	26	45
Liczba nagrań	38	52	90
Średnia wieku	$67,31 \pm 5,23$	$66,96 \pm 8,59$	$67,11 \pm 7,36$
Liczba kobiet	10	7	17
Liczba mężczyzn	9	19	28

4.1.3. Baza danych w języku hiszpańskim

Hiszpańska baza danych (PC-GITA) [37] zawiera nagrania mowy 50 pacjentów z PD i ich tyle samo osób zdrowych, dopasowanych pod względem wieku i płci. Wszyscy uczestnicy posługują się kolumbijską odmianą języka hiszpańskiego, a nagrania zostały zebrane zgodnie z protokołem uwzględniającym wymagania techniczne oraz zalecenia ekspertów z dziedzin lingwistyki, foniatrii i neurologii. Kolekcja obejmuje takie zadania jak podtrzymywane wymawianie samogłosek, ocenę diadokokinetyczną, 45 słów, 10 zdań, tekst do czytania i monolog. Podobnie jak w przypadku pozostałych baz wykorzystano

jedynie nagrania samogłosek /a/, /e/, /i/, /o/ oraz /u/. Na jednego uczestnika przypadają trzy powtórzenia każdej z samogłosek. Szczegółowa charakterystyka zbioru została przedstawiona w tabeli 4.3

Tabela 4.3. Charakterystyka hiszpańskiej bazy danych

Kategoria	Osoby zdrowe (HC)	Osoby chore (PD)	Razem
Liczba osób	50	50	100
Liczba nagrań	150	150	300
Średnia wieku	$60,90 \pm 9,37$	$61,14 \pm 9,51$	$61,02 \pm 9,15$
Liczba kobiet	25	25	50
Liczba mężczyzn	25	25	50

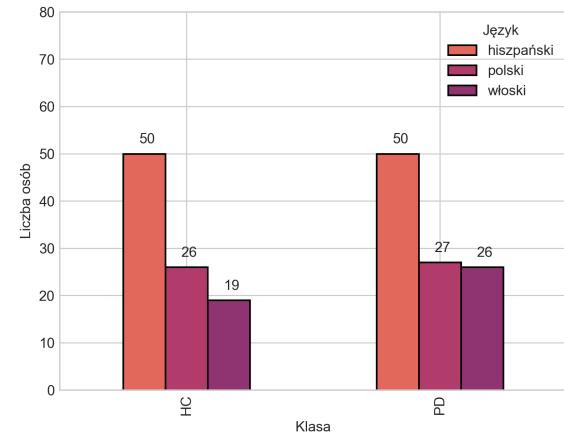
4.1.4. Podsumowanie wykorzystanych baz danych

Połączenie baz danych było możliwe ze względu na uniwersalność podstawy diagnostycznej, jaką są podtrzymywane samogłoski. Nagrania różniły się długością, jednak wykorzystano jedynie krótkie fragmenty (0,1 s), więc nie stanowiło to problemu. Wszystkie zostały zarejestrowane z częstotliwością próbkowania 44,1 kHz.

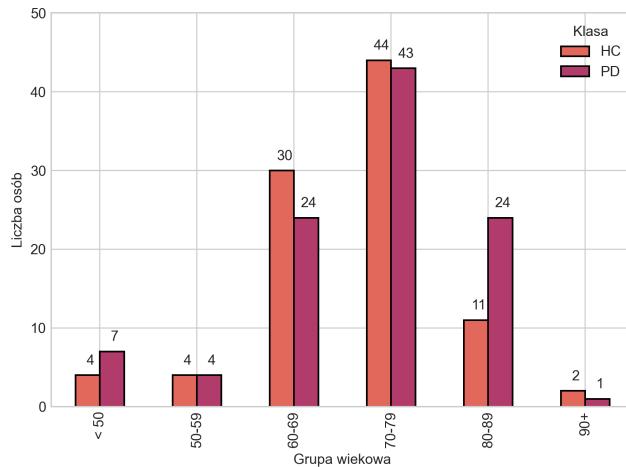
Trudno nie zauważyc znaczącej przewagi liczby nagrań w bazie hiszpańskojęzycznej (Rys. 4.2a). Jednak zdecydowano się na rezygnację z wyrównywania liczby nagrań w poszczególnych językach, z uwagi na to, że baza danych jest już stosunkowo niewielka, i dodatkowe ograniczenie liczby nagrań mogłoby znacznie zmniejszyć jej wartość i możliwości badawcze. Zachowano zbliżone proporcje wieku (różnica w średniej wieku mniejsza niż 5 lat) i płci pomiędzy grupą pacjentów a grupą porównawczą. Drobne różnice w liczbie próbek w poszczególnych przedziałach wiekowych nie powinny mieć istotnego wpływu na wyniki klasyfikacji. Liczba nagrań dla każdej samogłoski została wyrównana, tak by umożliwić obiektywne porównanie ich potencjalnego wykorzystania. Szczegółowe charakterystyki zbioru danych zostały przedstawione na Rys. 4.2.

Tabela 4.4. Charakterystyka stworzonej bazy danych

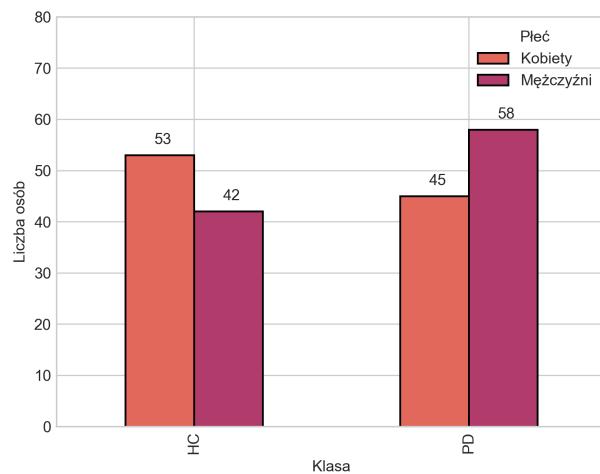
Kategoria	Osoby zdrowe (HC)	Osoby chore (PD)	Razem
Liczba osób	95	103	198
Liczba nagrań	255	229	484
Średnia wieku	$62,18 \pm 8,70$	$62,84 \pm 9,45$	$62,32 \pm 8,99$
Liczba kobiet	53	45	98
Liczba mężczyzn	42	58	100



(a) Udział nagrani w poszczególnych językach w ostatecznej bazie danych



(b) Rozkład klas w grupach wiekowych



(c) Rozkład płci w klasach

Rys. 4.2. Charakterystyka wykorzystanej bazy danych

4.2. Parametryzacja sygnału akustycznego

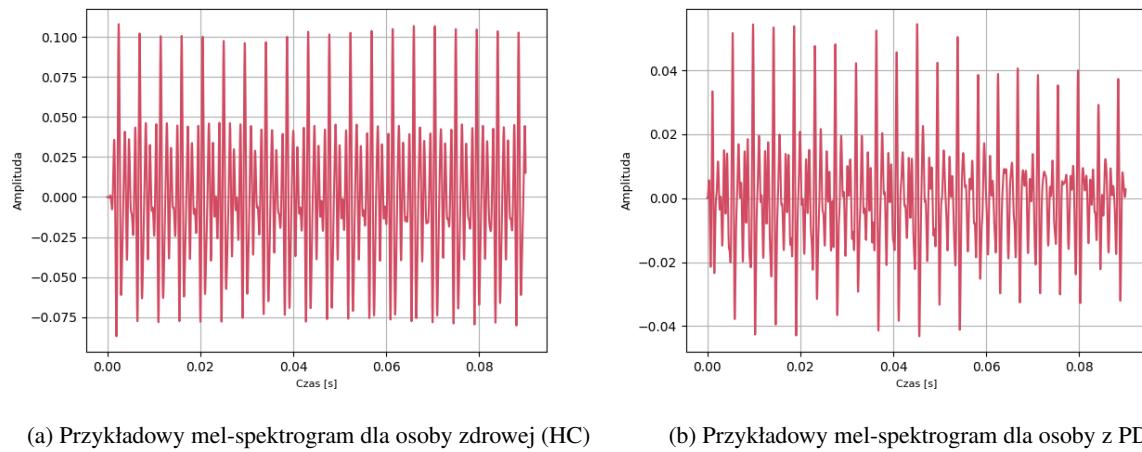
W kontekście projektowania rozwiązań związanych z uczeniem maszynowym, jednym z kluczowych aspektów jest staranne przygotowanie nagrań. Proces ten jest niezwykle istotny, ponieważ obejmuje kilka kluczowych etapów, które mają ogromny wpływ na jakość i skuteczność modelu. Dwa główne etapy, które wymagają szczególnej uwagi, to preprocessing i ekstrakcja cech. Preprocessing odnosi się do przetwarzania wstępnych danych dźwiękowych przed ich podaniem modelowi uczenia maszynowego. W ramach tego procesu usuwane są zakłócenia, szумy i niepożądane artefakty z nagrań. Może to obejmować filtrację, normalizację głośności, usuwanie niepotrzebnych fragmentów dźwięku oraz inne techniki poprawiające jakość sygnału. Poprawne wykonanie etapu preprocessingu może znacząco wpływać na jakość wyników uzyskiwanych przez modele. Drugim istotnym etapem jest ekstrakcja cech. W ramach tego procesu z nagrań mowy wydobywane są istotne parametry lub cechy, które mogą być używane przez modele uczenia maszynowego do rozpoznawania wzorców. Wybór odpowiednich cech i ich dokładna ekstrakcja mają kluczowe znaczenie, ponieważ to od nich zależy, jakie informacje zostaną dostarczone modelowi do analizy.

4.2.1. Przygotowanie nagrań

Nagrania zostały przycięte, tak by nie zawierały początkowych i końcowych fragmentów ciszy. Ten proces eliminuje niepotrzebne fragmenty i skupia analizę jedynie na istotnych fragmentach mowy, co może poprawić skuteczność modelu. Usunięcie ciszy na początku i końcu nagrań jest szczególnie ważne, gdyż może to pomóc w uniknięciu zakłóceń lub nieistotnych informacji, które mogą wpływać na wyniki klasyfikacji. Skorzystano z pakietu *Librosa* do automatycznego usunięcia niepożądanych fragmentów, a następnie każde z nagrań zostało przeanalizowane w programie *Audacity*. Upewniono się, że nagrania zostały poprawnie przetworzone oraz wprowadzono ręcznie ewentualne poprawki.

Z uwagi na wykorzystanie różnych baz danych, charakteryzujących się zróżnicowanymi warunkami nagrywania, zdecydowano się na ograniczenie wpływu otoczenia na dokładność klasyfikacji. W tym celu zastosowano filtr pasmowoprzepustowy, który pozwolił na eliminację niepotrzebnych odstających częstotliwości, które mogą nie mieć znaczenia dla analizy mowy. Wykorzystanie tego filtra zapobiegło również jedynie dostosowaniu modelu do cech charakteryzujących odstające częstotliwości, takie jak chrypka czy inne zakłócenia dźwiękowe. W efekcie przekazywane były jedynie częstotliwości zawarte w przedziale między 500 Hz a 1500 Hz.

Wykorzystano bardzo krótkie fragmenty nagrań o długości 0,1 s. Sygnały po przekształceniach przedstawiono na rysunku 4.3.



(a) Przykładowy mel-spektrogram dla osoby zdrowej (HC)

(b) Przykładowy mel-spektrogram dla osoby z PD

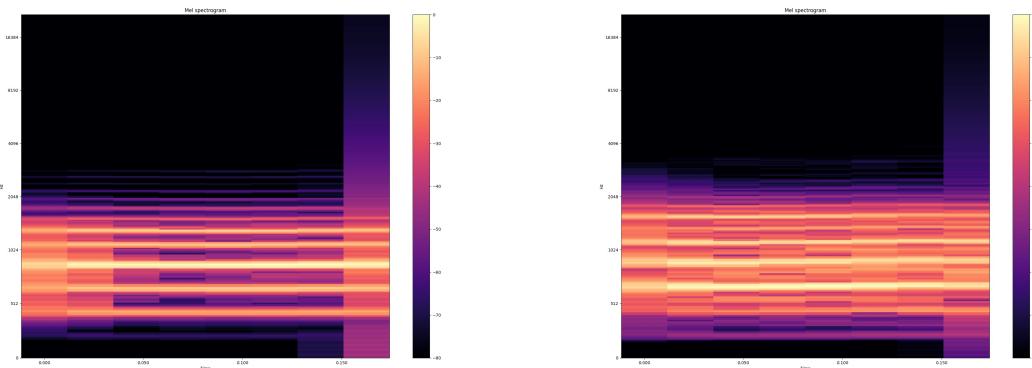
Rys. 4.3. Porównanie przykładowego nagrania osoby zdrowej (HC) oraz chorej (PD), po preprocessingu (hiszpańskojęzyczny zbiór danych, samogłoska /a/)

4.2.2. Obliczanie mel-spektrogramów

Zdecydowano się na przedstawienie przygotowanych nagrań w formie graficznej w postaci mel-spektrogramów, które zostaną następnie użyte do dalszej analizy. Spektrogramy to graficzna reprezentacja zmian widma dźwięku w funkcji czasu. Wykorzystując transformację Fouriera, sygnał dźwiękowy jest podzielony na krótkie ramki czasowe, a następnie dla każdej ramki obliczane jest widmo amplitudowe. Spektrogram przedstawia te widma w formie kolorowej mapy, gdzie oś pozioma reprezentuje czas, oś pionowa reprezentuje częstotliwość, a intensywność koloru oznacza amplitudę. Spektrogramy są często wykorzystywane w analizie dźwięku i sygnałów akustycznych. Pozwalają one zobaczyć, jak zmienia się struktura częstotliwościowa dźwięku w czasie, co może być przydatne do analizy cech akustycznych, identyfikacji dźwięków, diagnostyki medycznej czy rozpoznawania mowy. Dzięki spektrogramom można wizualnie analizować zmiany w widmie dźwięku, takie jak formanty samogłoskowe, obecność szumów czy artefaktów.

Natomiast mel-spektrogramy to wariant spektrogramów, w których oś częstotliwości jest przekształcona na skalę Mel, czyli nieliniową skalę częstotliwości, która bardziej odpowiada percepceji ludzkiego słuchu. Stosuje się ją, aby lepiej odzwierciedlić sposób, w jaki odbieramy i rozpoznajemy różnice między różnymi tonami dźwięków. Przez uwzględnienie charakterystyk percepceji słuchowej człowieka, mel-spektrogramy dostarczają bardziej reprezentatywnych danych akustycznych dla analizy i klasyfikacji dźwięku. Proces obliczeń mel-spektrogramów został zrealizowany przy użyciu biblioteki Pythona, Librosa. Wygenerowane mel-spektrogramy zostały przedstawione jako obrazy w skali logarytmicznej (Rys. 4.4).

Ważnym aspektem przy wyznaczaniu mel-spektrogramów są parametry ekstrakcji cech takie jak liczba pasm mel, rozmiar ramki analizy, długość przesunięcia (ang. *overlap*) oraz rozdzielcość czasowa i częstotliwościowa. Różne kombinacje tych parametrów mogą prowadzić do różnych szczegółów w



(a) Przykładowy mel-spektrogram dla osoby zdrowej

(b) Przykładowy mel-spektrogram dla osoby z PD

Rys. 4.4. Porównanie mel-spektrogramów osoby z PD oraz osoby zdrowej (hiszpańskojęzyczny zbiór danych, samogłoska /a/)

reprezentacji dźwięku. Wybór tych wartości został przeprowadzony na podstawie dostępnej literatury oraz własne eksperymenty.

Rozmiar okna czasowego jest jednym z kluczowych parametrów obliczeń melspekrogramów. W kontekście diagnozowania choroby Parkinsona istnieje potrzeba uwzględnienia zarówno krótkotrwałych, dynamicznych zmian w mowie, jak i dłuższych charakterystyk. Wybór okna o długości 2048 zostało dokonany w celu uzyskania odpowiedniej rozdzielczości w dziedzinie czasu i częstotliwości, co pozwala na dokładniejsze odwzorowanie tych cech w melspekrogramie. Długość przesunięcia określa, co ile próbek dźwiękowych nowe okno czasowe jest nakładane na sygnał. W celu zachowania równowagi między dokładnością w dziedzinie czasu a efektywnością obliczeń wybrano *hop_length* o wartości 512. Jest to kompromis, który pozwala na zachowanie informacji o krótkotrwałych zmianach w sygnale, co jest istotne w diagnozowaniu choroby Parkinsona. Liczba pasm Mel określa, ile pasm częstotliwościowych zostanie wygenerowanych w mel-spektrogramie. W celu uzyskania szczegółowej reprezentacji widma dźwiękowego wybrano wartość równą 320. Choroba Parkinsona może wprowadzać subtelne zmiany w mowie, dlatego ważne jest uwzględnienie wysokiej rozdzielczości w dziedzinie częstotliwości, co pozwala na dokładniejszą analizę sygnału mowy.

Często praktyką przed przekazaniem danych do sieci neuronowych (CNN) jest ich skalowanie. Pomaga ono zapewnić zbieżność, stabilność i efektywność procesu uczenia się modelu, nie wprowadzając dodatkowych informacji. Dlatego zdecydowano się na skalowanie metodą *Min-Max* na zakres od 0 do 1 opisaną wzorem (4.1).

$$\text{Skalowanie Min-Max: } X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (4.1)$$

4.2.3. Augmentacja

Aktualny stan wiedzy w dziedzinie automatycznej diagnostyki choroby Parkinsona ukazuje, że konwolucyjne sieci neuronowe (CNN) znaczco poprawiły wyniki w zadaniach przetwarzania mowy. Jednakże, aby uniknąć przeuczenia, konieczna jest duża ilość danych treningowych. Dostępne bazy danych zawierają zwykle do kilkuset nagrań, co ogranicza możliwość uwzględnienia wielu istotnych cech i stworzenia stabilnego modelu klasyfikacyjnego.

W początkowych eksperymentach przeprowadzonych w ramach tej pracy nie udało się osiągnąć oczekiwanych wyników, ze względu na bardzo duże przeuczenie (ang. *overfitting*) testowanych modeli. Jest to częsty problem, który można rozwiązać dzięki praktyce znanej jako augmentacja danych. Polega ona na modyfikacji oryginalnych próbek, co prowadzi do zwiększenia ilości danych treningowych. Augmentacja danych znaczco poprawia zdolność modeli CNN do generalizacji i zwiększa ich odporność na przeuczenie.

W badaniach opisanych w publikacji [38] zastosowano sześć technik augmentacji danych specyficznych dla mowy w celu polepszenia zdolności modelu do generalizacji na dane, które nie były wcześniej widziane. Wyniki tych badań pokazują, że techniki augmentacji danych specyficznych dla mowy istotnie poprawiają zdolność do wykrywania zaburzeń mowy u pacjentów z chorobą Parkinsona. Podobne znaczenie augmentacji danych jako istotnego czynnika wpływającego na skuteczność rozwiązań w dziedzinie klasyfikacji choroby Parkinsona podkreślono również w publikacji [30].

Dlatego w niniejszej pracy zdecydowano się zastosować 4 techniki augmentacyjne: przesunięcie w czasie, spowolnienie, przyspieszenie oraz losową zmianę wysokości dźwięku. Wszystkie modyfikacje zostały przeprowadzone na sygnałach przed wyznaczeniem mel-spektrogramów. Wpływ zmian na mel-spektrogramy przedstawiono na Rys 4.5.

- a) Przesunięcie w czasie (ang. *time shifting*): Aby zapewnić, że model nie dostosowuje się do lokalizacji czasowej danej próbki mowy lub głosu, zmieniana jest kolejność sygnału. Jest to osiągane poprzez przesunięcie sygnałów w prawo wzdłuż osi czasu o losową ilość, która jest mniejsza niż długość wejściowego nagrania audio.
- b) Losowa zmiana wysokości dźwięku (ang. *pitch change*): Składowe częstotliwości próbek są losowo przesuwane w dół lub w górę, przy czym należy zadbać o to, aby długość nie uległa zmianie poprzez rozciągnięcie oryginalnej próbki o losową ilość czasu z przedziału [3; 5] w dziedzinie czasu, a następnie jej resampling.
- c) Spowolnienie (ang. *slow-down*): Przez rozciągnięcie w czasie próbki dźwiękowej zapewniamy, że model sieci neuronowej nie dostosowuje się tylko do prędkości mowy lub głosu badanego podmiotu. Współczynnik spowolnienia jest losowo wybierany z przedziału [0,2; 0,8]. Jest to osiągane poprzez resampling oryginalnej próbki.

- d) Przyspieszenie (ang. *speed-up*): Tak samo, jak spowolnienie dźwięku, losowe przyspieszenie ma na celu zapobieżenie dostosowywaniu modelu do prędkości mowy mówcy. Współczynnik przyspieszenia jest losowo wybierany z przedziału [1,2; 2,5]. Jest to osiągane poprzez resampling oryginalnej próbki.

4.3. Metody klasyfikacji

Zdecydowano się na graficzne reprezentacje sygnałów, co wymaga też doboru odpowiednich metod klasyfikacji. Popularnym rozwiążaniem są konwolucyjne sieci neuronowe (ang. *Convolutional Neural Network*, CNN). Jest to rodzaj zaawansowanych modeli uczenia maszynowego, które zostały pierwotnie zaprojektowane do przetwarzania i analizy danych wizualnych, takich jak obrazy i filmy. Jednakże, ze względu na swoją skuteczność w ekstrakcji cech z danych przestrzennych, CNN znalazły zastosowanie także w dziedzinach związanych z przetwarzaniem dźwięku, w tym w analizie mel-spektrogramów i rozpoznawaniu mowy.

Główną cechą konwolucyjnych sieci neuronowych są warstwy konwolucyjne. CNN wykorzystują warstwy konwolucyjne do przetwarzania danych wejściowych. Te warstwy wykorzystują filtry konwolucyjne, które przesuwają się po wejściowych danych, wyodrębniając lokalne wzorce, takie jak krawędzie, tekstury i inne cechy. To pozwala na automatyczną ekstrakcję istotnych cech. CNN często zawierają też warstwy pooling, które zmniejszają rozmiar danych wyjściowych z warstw konwolucyjnych. Pooling pomaga zmniejszyć liczbę parametrów w sieci i sprawia, że sieć jest bardziej odporna na przetwarzanie danych o różnych rozmiarach. CNN zwykle składają się z wielu warstw konwolucyjnych i warstw fully connected, które uczą się reprezentacji coraz bardziej abstrakcyjnych cech danych. Dzięki tym cechom są wysoce zdolne do rozpoznawania wzorców w danych, co czyni je skutecznymi w zadaniach klasyfikacji, detekcji obiektów, rozpoznawaniu mowy i wielu innych. Dużą zaletą sieci konwolucyjnych jest możliwość skorzystania z uczenia z transferem, co oznacza wykorzystanie pre-trenowanych modeli na dużych zbiorach danych, a następnie dostosowanie ich do konkretnej dziedziny lub zadania.

Podsumowując, konwolucyjne sieci neuronowe (CNN) są potężnym narzędziem w analizie graficznych reprezentacji danych i są powszechnie stosowane w wielu dziedzinach, aby automatycznie ekstrahować i analizować cechy złożonych danych przestrzennych. Dzięki ich zdolności do hierarchicznego uczenia się i rozpoznawania wzorców, są używane w różnych zastosowaniach, w tym również w medycynie. Dlatego zdecydowano się wykorzystać tę metodę do klasyfikacji mel-spektrogramów.

4.3.1. Wykorzystane architektury

Ze względu na ograniczoną dostępność danych głosowych pacjentów z chorobą Parkinsona, zastosowano technikę znaną jako *transfer learning*. Transfer learning umożliwia wykorzystanie wcześniej wytrenowanych modeli, które zostały nauczane na dużym zbiorze danych, do zadań diagnostycznych z wykorzystaniem ograniczonej ilości dostępnych danych głosowych.

Kluczowym krokiem w transfer learningu było wybranie modeli, które miały dostępne wagie wytrenowane na dużym i takim samym zbiorze danych. Zdecydowano się na ImageNet, ponieważ jest to jedna z najważniejszych baz danych obrazów wykorzystywana w dziedzinie komputerowego widzenia i uczenia maszynowego. Wybrano modele takie jak Xception, MobileNetV2, InceptionV3, VGG16 i ResNet50, co było uzasadnione dostępnością wstępnie wytrenowanych wag, które zawierały ogólne cechy obrazów. To pozwoliło na zaoszczędzenie czasu i zasobów, które byłyby potrzebne do wytrenowania sieci od podstaw na niewielkim zbiorze danych.

W ramach pracy magisterskiej skoncentrowano się na porównaniu różnych architektur w kontekście automatycznej diagnostyki choroby Parkinsona na podstawie analizy głosu.

- a) VGG16: To przykład klasycznego modelu CNN o głębokiej architekturze, opracowany przez Visual Geometry Group na Uniwersytecie Oksfordzkim. Jest znany ze swojej prostoty i skuteczności w ekstrakcji cech z obrazów. Pomimo swojej głębokości, może być użyteczny w diagnostyce na podstawie głosu, szczególnie jeśli głosowe dane wejściowe są w formie spektrogramów.
- b) ResNet50: Jest to znana architektura CNN, która wprowadza innowacyjny pomysł na połączenia pomostowe, eliminujące problem znikającego gradientu w głębokich sieciach. Dzięki temu ResNet50 jest w stanie efektywnie uczyć się reprezentacji danych. W diagnostyce na podstawie głosu może pomóc w wykrywaniu subtelnych cech charakterystycznych dla choroby Parkinsona.
- c) Xception: Jest to model głębokiej nauki opracowany przez François Chollet w 2016 roku. Jest to jedna z odmian architektury konwolucyjnej sieci neuronowej (CNN), która wyróżnia się wyjątkową zdolnością do wykrywania cech hierarchicznych w obrazach. W kontekście diagnostyki choroby Parkinsona na podstawie głosu Xception może pomóc w wyodrębnianiu istotnych cech z obrazów spektrogramów głosowych.
- d) InceptionV3: Charakteryzuje się wykorzystaniem tzw. *inception modules*, które pozwalają na efektywne wykrywanie wielu skal i rodzajów cech w obrazach. To sprawia, że InceptionV3 może być przydatny w diagnostyce choroby Parkinsona, gdzie istotne mogą być różne aspekty głosu. Został zaprojektowany przez Google.
- e) MobileNetV2: Jest to przykład lekkiej i efektywnej architektury CNN, zaprojektowanej z myślą o urządzeniach mobilnych. Jej cechą charakterystyczną jest niska ilość parametrów i małe obliczenia, co sprawia, że jest idealna do zastosowań w zasobochłonnych zadaniach takich jak analiza głosu na urządzeniach mobilnych.

Wybór tych konkretnych klasyfikatorów wynikał z różnorodności ich architektur i specjalizacji. Choroba Parkinsona manifestuje się na różne sposoby w głosie pacjentów, dlatego zróżnicowane modele w różny sposób mogą wspomagać wydobycie istotnych cech. Xception i InceptionV3 są znane z efektywności w ekstrakcji wielu rodzajów cech, co jest istotne przy analizie złożonych danych głosowych. MobileNetV2, z kolei, jest lekki i mobilny, co pozwala na przenośność rozwiązania do urządzeń mobilnych,

które mogą być używane w codziennym monitoringu pacjentów. VGG16 i ResNet50, choć głębokie, mają zdolność do wyodrębniania skomplikowanych wzorców, co może być przydatne w diagnozie. Wybór tych różnorodnych klasyfikatorów pozwoli na rzetelne porównanie ich wydajności i skuteczności w diagnostyce na podstawie głosu, co może przyczynić się do lepszego zrozumienia, który model najlepiej radzi sobie z tym zadaniem.

Po wstępny treningu na ImageNet przeprowadzono proces fine-tuningu na zbiorze danych zawierającym nagrania głosowe pacjentów związane z diagnozą choroby Parkinsona. Dostrajanie jest klużowym etapem w procesie dostosowywania wstępnie wytrenowanych klasyfikatorów do konkretnej diagnozy na podstawie głosu. W trakcie fine-tuningu modele były dostosowywane, aby nauczyć się rozpoznawania specyficznych cech akustycznych charakterystycznych dla pacjentów z chorobą Parkinsona. Umożliwiło to dostosowanie wag i parametrów modeli do konkretnego zadania diagnostycznego. W ten sposób klasyfikatory stały się bardziej odpowiednie do analizy dźwięku i diagnozy choroby Parkinsona na podstawie danych głosowych.

4.3.2. Dobór parametrów i optymalizacja modeli

Wszystkie eksperymenty przeprowadzono przy użyciu tych samych parametrów. Testowano różne kombinacje, jednak te ustawienia okazały się optymalne dla badanego problemu. W eksperimentach uwzględniono takie aspekty, jak rozmiar partii (ang. *batch-size*), funkcję straty i optymalizator.

Wybór rozmiaru partii, czyli liczby próbek przetwarzanych jednocześnie przez sieć neuronową, ma znaczenie dla równowagi pomiędzy wydajnością obliczeniową a stabilnością procesu uczenia. Przyjęto rozmiar partii równy 32, co jest kompromisem między efektywnym wykorzystaniem zasobów obliczeniowych a stabilnością uczenia.

Funkcja straty, czyli binarna entropia krzyżowa, została wybrana ze względu na swoją skuteczność w problemach klasyfikacji binarnej. Minimalizacja funkcji straty pozwala modelowi lepiej dopasować się do rzeczywistych klas i ocenić jakość predykcji sieci. Ta metryka jest często stosowana w zadaniach klasyfikacji, ponieważ skupia się na porównywaniu prognozowanych klas z rzeczywistymi wartościami.

Wybór optymalizatora SGD (ang. *Stochastic Gradient Descent*) był wynikiem eksperimentalnej analizy różnych algorytmów optymalizacji i ich właściwości. Ponadto jest on najczęściej spotykany w literaturze dotyczącej tego problemu badawczego. W każdej iteracji SGD losowo wybiera małe losowe podzbiory danych treningowych (mini-pakiety) i oblicza gradient funkcji kosztu tylko na podstawie tych próbek, co pozwala na szybszą aktualizację wag modelu. Ta stochastyczna natura SGD pomaga uniknąć utknięcia w minimach lokalnych.

Bardzo ważny w tym problemie klasyfikacyjnym jest odpowiedni dobór współczynnika uczenia (ang. *learning rate*). Wartość tego parametru wpływa na tempo, z jakim model aktualizuje swoje wagi podczas procesu uczenia. Wysoki learning rate może prowadzić do oscylacji i trudności w zbieżności algorytmu, podczas gdy zbyt niski może spowolnić proces uczenia. Dlatego optymalny wybór learning rate jest istotnym krokiem w dostrojeniu algorytmu SGD do konkretnego problemu. Zdecydowano się na ustawienie wartości współczynnika uczenia na 0,0005.

Podczas treningu modelu użyto również monitorowania funkcji straty. Mechanizm ten umożliwia wczesne zatrzymanie treningu, jeśli funkcja straty przestaje się poprawiać przez określoną liczbę epok (przyjęto 5).

4.3.3. Podział danych

Klasyczną techniką w uczeniu maszynowym jest podział zbioru danych na 3 podzbiory: treningowy, testowy oraz walidacyjny. W problemie klasyfikacji celem jest stworzenie algorytmu, który na podstawie znanych sobie, opisanych wzorców, będzie w stanie efektywnie rozpoznawać wzorce nieopisane i dotąd sobie nieznane. Dlatego wydziela się zbiór treningowy, na podstawie, którego model uczy się wzorców i dopasowuje do danych. Aby uniknąć nadmiernego dopasowania danych do modelu (ang. overfitting) stosuje się również zbiór walidacyjny, który przeprowadza bezstronną ocenę dopasowania modelu do zbioru danych szkoleniowych podczas dostrajania hiperparametrów modelu. Na koniec ocenia się skuteczność modelu na niezależnym zbiorze testowym.

Innym często wykorzystywana techniką jest wykorzystanie krosswalidacji (ang. *cross-validation*). Jest to technika oceny wydajności modelu maszynowego, która polega na podziale dostępnych danych na kilka części, z których jedna służy jako zbiór testowy, a pozostałe jako zbiór treningowy. Proces ten jest wielokrotnie powtarzany, aby uzyskać stabilniejszą ocenę wydajności modelu i uniknąć przeuczenia. Najpopularniejszym typem krosswalidacji jest walidacja krzyżowa typu k-fold, gdzie dane są dzielone na k podzbiorów, a każdy z nich jest używany jako zbiór testowy w jednym z k cykli walidacji.

W pracy zastosowano 10-krotną walidację krzyżową oraz przyjęto podział na zbiór treningowy i testowy według proporcji 80:20. Ważnym aspektem jest fakt, że podział nagrań dla różnych samogłosek przeprowadzono biorąc pod uwagę tożsamość mówcy (podejście *subject-wise*). To znaczy, że na przykład w zbiorze treningowym samogłoski /a/ oraz /e/ znajdują się nagrania pochodzące od tych samych osób mówiących. Ta zasada obowiązuje we wszystkich podzbiorach. Pomaga to zachować spójne warunki uczenia i obiektywnie porównać otrzymane modele. Nie dopuszcza to też możliwości, że model uczy się wzorców związanych z osobą, a nie chorobą Parkinsona.

4.4. Metody ewaluacji wyników

W celu oceny skuteczności różnych metod klasyfikacji choroby Parkinsona przeprowadzono badania wykorzystujące różnorodne techniki ewaluacji. Niniejsza sekcja zawiera prezentację wybranych metod, które zostały użyte w ramach przeprowadzonych eksperymentów. Kluczowym aspektem tych eksperymentów jest konieczność zastosowania odpowiednich technik walidacji wyników. Techniki te nie tylko umożliwiają obiektywną ocenę efektywności proponowanych rozwiązań, ale także pozwalają na porównanie ich ze sobą. Dzięki temu pozwalają lepiej zrozumieć, które podejścia są najbardziej obiecujące w kontekście diagnozowania choroby Parkinsona i polepszania dokładności klasyfikacji, co jest kluczowe dla postępu w dziedzinie medycyny i diagnostyki tej choroby.

4.4.1. Krzywe uczenia

Krzywe uczenia są skutecznym narzędziem do wizualizacji procesu uczenia modelu. W trakcie eksperymentów prowadzono monitorowanie dwóch kluczowych krzywych: dokładność (ang. *accuracy*) i funkcję kosztu (ang. *loss*). Krzywa dokładności przedstawia zmiany dokładności modelu w trakcie treningu i walidacji, podczas gdy krzywa funkcji kosztu pokazuje, jak zmienia się funkcja kosztu w trakcie uczenia. Analiza tych krzywych dostarcza istotnych informacji na temat efektywności uczenia modelu. Pozwala ocenić, czy model uczy się poprawnie, czy może występują problemy z nadmiernym dopasowaniem (ang. *overfitting*) lub niedostatecznym dopasowaniem (ang. *underfitting*). Dzięki tym analizom można dokonać wniosków na temat jakości modelu i ewentualnie dostosować hiperparametry lub strategię treningową w celu uzyskania lepszych wyników. Wartościowe informacje o procesie uczenia są kluczowe dla udanej implementacji modelu i oceny jego skuteczności.

4.4.2. Metryki walidacyjne

Podczas oceny skuteczności klasyfikacji, korzystano z najbardziej popularnych metryk, które pozwalają ocenić jakość predykcji modelu. W podanych wzorach zastosowano oznaczenia: TP – True Positive, TN – True Negative, FP – False Positive, FN – False Negative.

a) Dokładność (ang. *accuracy*)

To prosta i intuicyjna miara, która oblicza stosunek poprawnie sklasyfikowanych próbek do wszystkich próbek. Wyraża ona ogólną skuteczność klasyfikacji. Wyrażana jest wzorem 4.2

$$\frac{TP + TN}{TP + FN + TN + FP} \quad (4.2)$$

b) Precyzja (ang. *precision*)

Mierzy stosunek prawdziwie pozytywnych predykcji do sumy prawdziwie pozytywnych i fałszywie pozytywnych predykcji. Wyraża zdolność modelu do identyfikowania prawdziwie pozytywnych przypadków. Określa się ją wzorem 4.3

$$\frac{TP}{TP + FP} \quad (4.3)$$

c) Czułość (ang. *recall*)

Jest to stosunek prawdziwie pozytywnych predykcji do sumy prawdziwie pozytywnych i fałszywie negatywnych predykcji. Wyrażana jest wzorem 4.4 i określa zdolność modelu do wykrywania wszystkich prawdziwie pozytywnych przypadków.

$$\frac{TP}{TP + FN} \quad (4.4)$$

d) Miara F1 (ang. *F1-score*)

To harmoniczna średnia precyzyji i czułości. Jest używana jako miara równowagi między precyzyją a czułością. Wyższe wartości wskazują na lepszą jakość klasyfikacji modelu. Wyraża się ją wzorem 4.5.

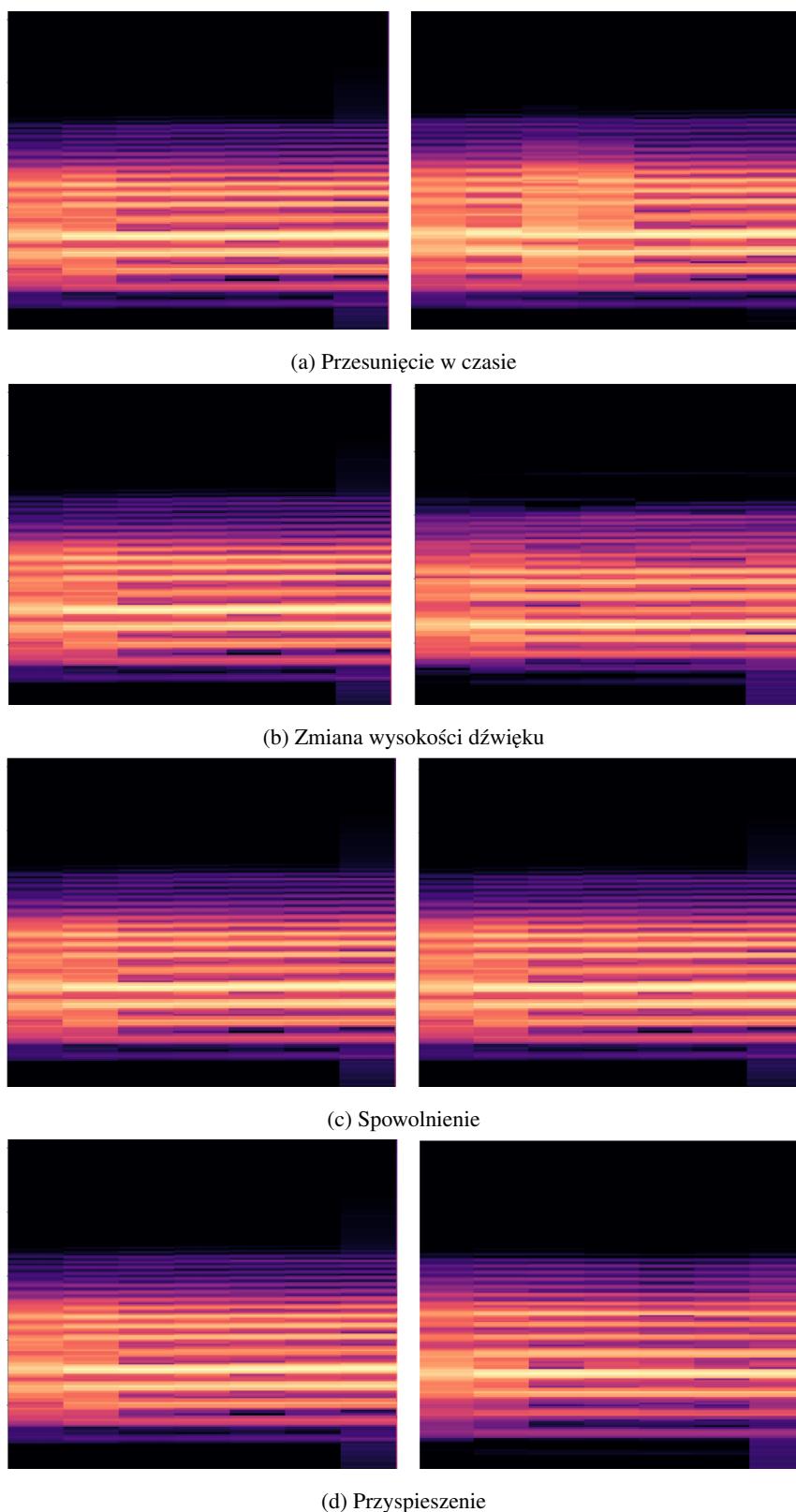
$$\frac{2 * TP}{2 * TP + FP + FN} \quad (4.5)$$

4.4.3. Macierz pomyłek

Macierz pomyłek to narzędzie służące do analizy wyników klasyfikacji. Przedstawia ona liczbę prawidłowo i błędnie zaklasyfikowanych próbek dla każdej klasy. Na podstawie macierzy pomyłek możemy obliczyć błędy I i II rzędu.

- a) Błąd I rzędu (ang. *False Positive*) oznacza błędную klasyfikację przypadku, który jest negatywny, jako pozytywny. W kontekście Parkinsona, błąd I rzędu oznacza, że model zaklasyfikował zdrową osobę jako chorą na chorobę Parkinsona.
- b) Błąd II rzędu (ang. *False Negative*) oznacza błędную klasyfikację przypadku, który jest pozytywny, jako negatywny. W kontekście Parkinsona, błąd II rzędu oznacza, że model zaklasyfikował osobę chorą na chorobę Parkinsona jako zdrową.

Analiza błędów I i II rzędu ma duże znaczenie w kontekście ewaluacji systemów diagnostycznych. Błąd I rzędu może prowadzić do niezdiagnozowania choroby u pacjenta, podczas gdy błąd II rzędu może prowadzić do błędnego zdiagnozowania osoby zdrowej jako cierpiącej na chorobę Parkinsona. Ważne jest, aby oceniać i minimalizować te błędy w celu uzyskania jak najdokładniejszej klasyfikacji.



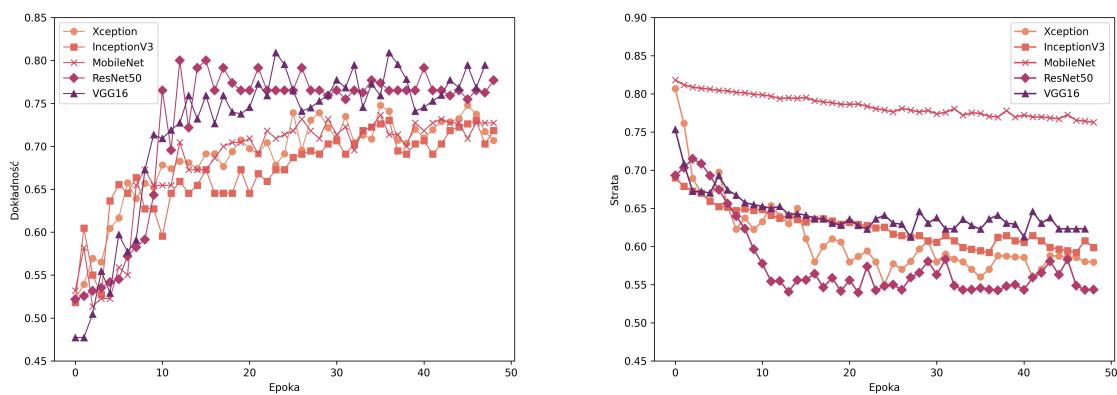
Rys. 4.5. Porównanie różnych technik augmentacji i ich wpływu na wygląd spektrogramu (zbiór danych PC-GITA, samogłoska /a/, HC)

5. Wyniki badań

5.1. Samogłoska /a/

Tabela 5.1. Wyniki otrzymane dla samogłoski /a/

Model	VGG16	Resnet50	Xception	InceptionV3	MobileNetV2
Accuracy	0.751	0.724	0.672	0.692	0.714
Precision	0.752	0.703	0.669	0.721	0.712
Recall	0.751	0.713	0.663	0.689	0.709
F1-score	0.749	0.727	0.675	0.688	0.711
Loss	0.645	0.563	0.611	0.676	0.750

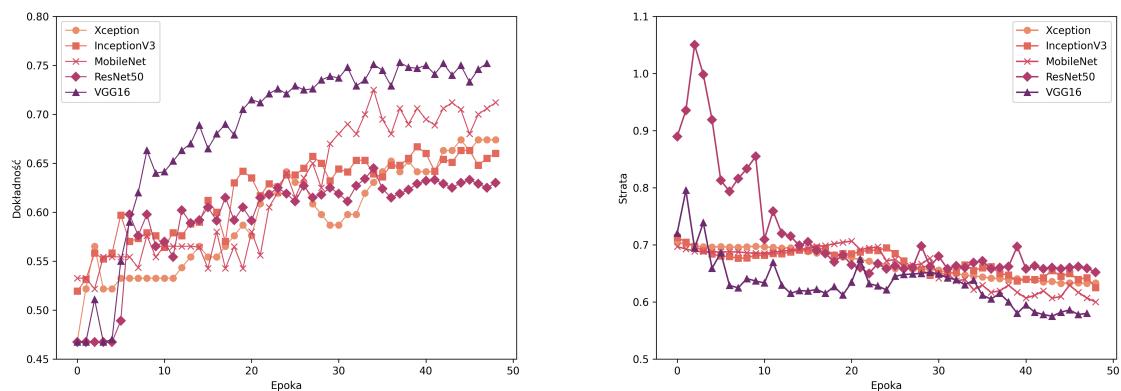


(a) Krzywe dokładności na zbiorze walidacyjnym podczas uczenia (b) Krzywe funkcji straty na zbiorze walidacyjnym podczas uczenia

Rys. 5.1. Porównanie krzywych uczenia dla różnych klasyfikatorów na zbiorze walidacyjnym (samogłoska /a/)

Tabela 5.2. Wyniki otrzymane dla samogłoski /e/

Model	VGG16	Resnet50	Xception	InceptionV3	MobileNetV2
Accuracy	0.733	0.621	0.653	0.644	0.700
Precision	0.729	0.625	0.640	0.641	0.721
Recall	0.733	0.620	0.633	0.638	0.689
F1-score	0.735	0.643	0.649	0.640	0.692
Loss	0.582	0.659	0.632	0.634	0.601



(a) Krzywe dokładności na zbiorze walidacyjnym podczas uczenia (b) Krzywe funkcji straty na zbiorze walidacyjnym podczas uczenia

Rys. 5.2. Porównanie krzywych uczenia dla różnych klasyfikatorów na zbiorze walidacyjnym (samogłoska /e/)

5.2. Samogłoska /e/

5.3. Samogłoska /i/

5.4. Samogłoska /o/

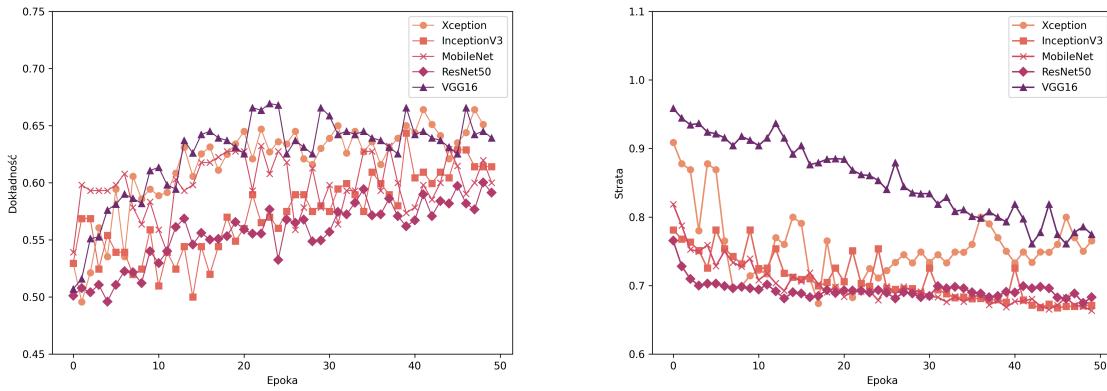
5.5. Samogłoska /u/

5.6. Połączenie samogłosek /a/, /e/, /i/, /o/, /u/

5.7. Zbiorcze podsumowanie wyników

Tabela 5.3. Wyniki otrzymane dla samogłoski /i/

Model	VGG16	Resnet50	Xception	InceptionV3	MobileNetV2
Accuracy	0.627	0.586	0.641	0.605	0.634
Precision	0.642	0.589	0.638	0.598	0.672
Recall	0.616	0.584	0.642	0.595	0.623
F1-score	0.605	0.572	0.643	0.597	0.614
Loss	0.775	0.699	0.751	0.686	0.656



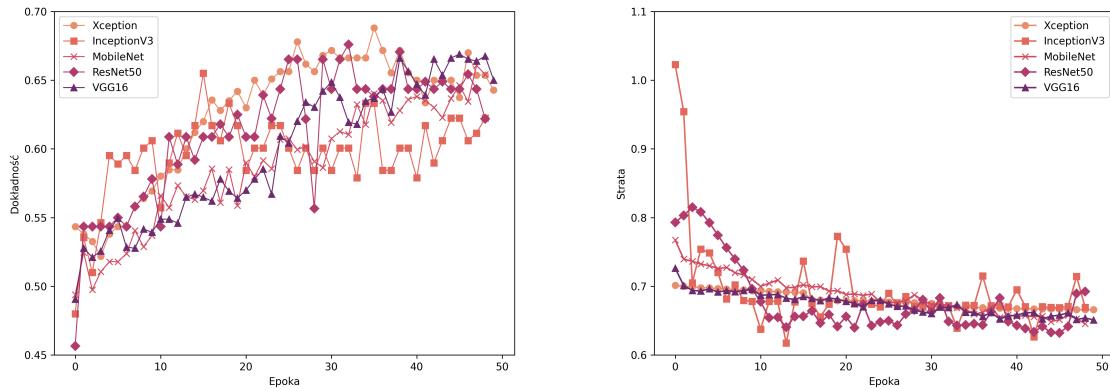
(a) Krzywe dokładności na zbiorze walidacyjnym podczas uczenia (b) Krzywe funkcji straty na zbiorze walidacyjnym podczas uczenia

Rys. 5.3. Porównanie krzywych uczenia dla różnych klasyfikatorów na zbiorze walidacyjnym (samogłoska /i/)**Tabela 5.4.** Wyniki otrzymane dla samogłoski /o/

Model	VGG16	Resnet50	Xception	InceptionV3	MobileNetV2
Accuracy	0.653	0.632	0.653	0.611	0.611
Precision	0.654	0.651	0.651	0.612	0.624
Recall	0.649	0.635	0.649	0.609	0.606
F1-score	0.650	0.633	0.646	0.601	0.594
Loss	0.674	0.679	0.662	0.679	0.677

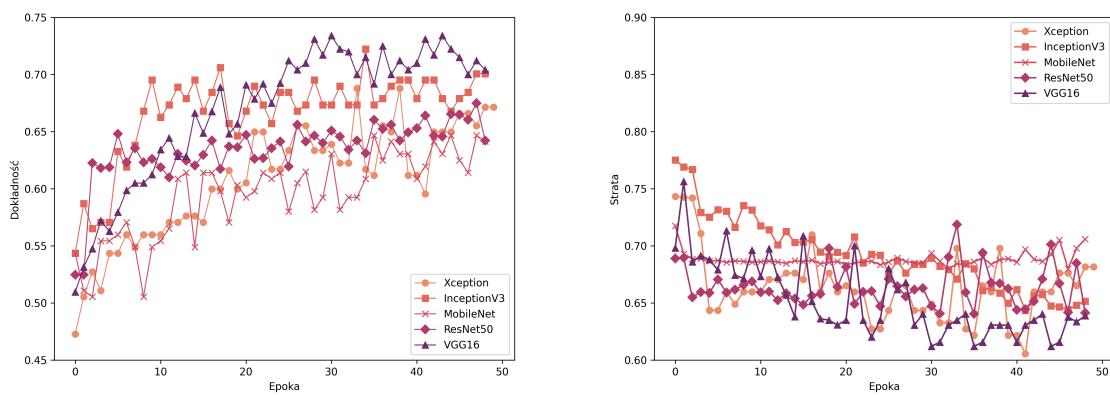
Tabela 5.5. Wyniki otrzymane dla samogłoski /u/

Model	VGG16	Resnet50	Xception	InceptionV3	MobileNetV2
Accuracy	0.724	0.642	0.663	0.705	0.642
Precision	0.730	0.689	0.682	0.714	0.686
Recall	0.715	0.636	0.657	0.709	0.635
F1-score	0.712	0.615	0.657	0.708	0.610
Loss	0.654	0.685	0.665	0.637	0.750



(a) Krzywe dokładności na zbiorze walidacyjnym podczas uczenia
 (b) Krzywe funkcji straty na zbiorze walidacyjnym podczas uczenia

Rys. 5.4. Porównanie krzywych uczenia dla różnych klasyfikatorów na zbiorze walidacyjnym (samogłoska /o/)



(a) Krzywe dokładności na zbiorze walidacyjnym podczas uczenia
 (b) Krzywe funkcji straty na zbiorze walidacyjnym podczas uczenia

Rys. 5.5. Porównanie krzywych uczenia dla różnych klasyfikatorów na zbiorze walidacyjnym (samogłoska /u/)

6. Analiza i interpretacja wyników

6.1. Samogłoski a informacja diagnostyczna

6.2. Architektury CNN a skuteczność klasyfikacji

7. Podsumowanie

Bibliografia

- [1] E. Dorsey i in.: „Global, regional, and national burden of Parkinson’s disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016”, w: *The Lancet Neurology* 17, paź. 2018, s. 939–953, DOI: [10.1016/S1474-4422\(18\)30295-3/](https://doi.org/10.1016/S1474-4422(18)30295-3)
- [2] World Health Organization: „Parkinson disease: a public health approach. Technical brief.”, w: 2022
- [3] National Institute on Aging: *Parkinson’s Disease Information*, <https://www.nia.nih.gov/health/parkinsons-disease>, U.S. Department of Health and Human Services, 2022
- [4] Iwona Nowakowska-Kempna Tatiana Lewicka Daniel Stompel: „Diagnoza i terapia zaburzeń mowy u pacjentów z chorobą Parkinsona”, w: *Logopedia Silesiana*, 3, 2009, s. 76–94
- [5] Mateusz Szurek: „Mowa pacjentki z chorobą Parkinsona w fazie zaawansowanej”, w: *Interdyscyplinarne Konteksty Pedagogiki Specjalnej* 14, 2018, 137–157, DOI: [10.14746/ikps.2016.14.08](https://doi.org/10.14746/ikps.2016.14.08)
- [6] „Zaburzenia głosu w chorobach neurologicznych”, w: *Forum Medycyny Rodzinnej* 13.6, 2019, s. 285 –291, ISSN: 1897-7839
- [7] Rania Khaskhoussy i Yassine Ben Ayed: „Improving Parkinson’s disease recognition through voice analysis using deep learning”, w: *Pattern Recognition Letters* 168, 2023, s. 64–70, ISSN: 0167-8655, DOI: <https://doi.org/10.1016/j.patrec.2023.03.011>
- [8] Jan Rusz i in.: „Guidelines for Speech Recording and Acoustic Analyses in Dysarthrias of Movement Disorders”, w: *Movement Disorders* 36.4, 2021, s. 803–814, DOI: <https://doi.org/10.1002/mds.28465>, eprint: <https://movementdisorders.onlinelibrary.wiley.com/doi/pdf/10.1002/mds.28465>
- [9] Javier Gamboa i in.: „Acoustic voice analysis in patients with Parkinson’s disease treated with dopaminergic drugs”, w: *Journal of Voice* 11.3, 1997, s. 314–320, ISSN: 0892-1997, DOI: [https://doi.org/10.1016/S0892-1997\(97\)80010-0](https://doi.org/10.1016/S0892-1997(97)80010-0)
- [10] Emilia Sitek i in.: „Ocena neuropsychologiczna i neuropsychiatryczna w chorobie Parkinsona – specyfika badania i dobór metod diagnostycznych”, w: *Polski Przegląd Neurologiczny* 9, list. 2013, s. 105–112, DOI: [10.5603/ppn.36264](https://doi.org/10.5603/ppn.36264)
- [11] C. Warren Olanow i Anthony H.V. Schapira: „Parkinson’s Disease”, w: *Harrison’s Principles of Internal Medicine*, 21e, red. Joseph Loscalzo i in., New York, NY: McGraw-Hill Education, 2022

- [12] Malco Rossi, Santiago Perez-Lloret i Marcelo Merello: „How much time is needed in clinical practice to reach a diagnosis of clinically established Parkinson’s disease?”, w: *Parkinsonism & Related Disorders* 92, 2021, s. 53–58, ISSN: 1353-8020, DOI: <https://doi.org/10.1016/j.parkreldis.2021.10.016>
- [13] Jie Mei, Christian Desrosiers i Johannes Frasnelli: „Machine Learning for the Diagnosis of Parkinson’s Disease: A Review of Literature”, w: *Frontiers in Aging Neuroscience* 13, 2021, ISSN: 1663-4365, DOI: [10.3389/fnagi.2021.633752](https://doi.org/10.3389/fnagi.2021.633752)
- [14] Manila Kodali, Sudarsana Reddy Kadiri i Paavo Alku: „Automatic classification of the severity level of Parkinson’s disease: A comparison of speaking tasks, features, and classifiers”, w: *Computer Speech & Language* 83, 2023, s. 101548, ISSN: 0885-2308, DOI: <https://doi.org/10.1016/j.csl.2023.101548>
- [15] T. Arias-Vergara i in.: „Unobtrusive Monitoring of Speech Impairments of Parkinson’S Disease Patients Through Mobile Devices”, w: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, s. 6004–6008, DOI: [10.1109/ICASSP.2018.8462332](https://doi.org/10.1109/ICASSP.2018.8462332)
- [16] Pawalai Kraipeerapun i Somkid Amornsamankul: „Using stacked generalization and complementary neural networks to predict Parkinson’s disease”, w: *2015 11th International Conference on Natural Computation (ICNC)*, 2015, s. 1290–1294, DOI: [10.1109/ICNC.2015.7378178](https://doi.org/10.1109/ICNC.2015.7378178)
- [17] Liaqat Ali i in.: „A Multi-model Framework for Evaluating Type of Speech Samples having Complementary Information about Parkinson’s Disease”, w: lip. 2019, s. 1–5, DOI: [10.1109/ICECCE47252.2019.8940696](https://doi.org/10.1109/ICECCE47252.2019.8940696)
- [18] Hariharan M, Kemal Polat i Sindhu a R: „A new hybrid intelligent system for accurate detection of Parkinson’s disease”, w: *Computer methods and programs in biomedicine* 113, sty. 2014, DOI: [10.1016/j.cmpb.2014.01.004](https://doi.org/10.1016/j.cmpb.2014.01.004)
- [19] Rahib Abiyev i Sanan Abizade: „Diagnosing Parkinson’s Diseases Using Fuzzy Neural System”, w: *Computational and Mathematical Methods in Medicine* 2016, sty. 2016, s. 1–9, DOI: [10.1155/2016/1267919](https://doi.org/10.1155/2016/1267919)
- [20] Liaqat Ali i in.: „Automated Detection of Parkinson’s Disease Based on Multiple Types of Sustained Phonations Using Linear Discriminant Analysis and Genetically Optimized Neural Network”, w: *IEEE Journal of Translational Engineering in Health and Medicine* PP, paź. 2019, s. 1–1, DOI: [10.1109/JTEHM.2019.2940900](https://doi.org/10.1109/JTEHM.2019.2940900)
- [21] Niousha Karimi Dastjerd i in.: „Fuzzy Classification Methods Based Diagnosis of Parkinson’s disease from Speech Test Cases”, w: *Current Aging Science* 12, czer. 2019, DOI: [10.2174/1874609812666190625140311](https://doi.org/10.2174/1874609812666190625140311)
- [22] Ewelina Majda-Zdanciewicz i in.: „Deep learning vs feature engineering in the assessment of voice signals for diagnosis in Parkinson’s disease”, w: *Bulletin of the Polish Academy of Sciences. Technical Sciences* 69.3, 2021, art. no. e137347, ISSN: 0239-7528

- [23] Hakan Gündüz i in.: „Deep Learning-Based Parkinson’s Disease Classification Using Vocal Feature Sets”, w: *IEEE Access* PP, sierp. 2019, s. 1–1, DOI: [10.1109/ACCESS.2019.2936564](https://doi.org/10.1109/ACCESS.2019.2936564)
- [24] Máté Hireš i in.: „Convolutional neural network ensemble for Parkinson’s disease detection from voice recordings”, w: *Computers in Biology and Medicine* 141, 2022, s. 105021, ISSN: 0010-4825, DOI: <https://doi.org/10.1016/j.combiomed.2021.105021>
- [25] Laiba Zahid i in.: „A Spectrogram-Based Deep Feature Assisted Computer-Aided Diagnostic System for Parkinson’s Disease”, w: *IEEE Access* 8, 2020, s. 35482–35495, DOI: [10.1109/ACCESS.2020.2974008](https://doi.org/10.1109/ACCESS.2020.2974008)
- [26] Renata Guatelli i in.: „Detection of Parkinson’s disease based on spectrograms of voice recordings and Extreme Learning Machine random weight neural networks”, w: *Engineering Applications of Artificial Intelligence* 125, 2023, s. 106700, ISSN: 0952-1976, DOI: <https://doi.org/10.1016/j.engappai.2023.106700>
- [27] E Gelvez-Almeida i in.: „Classification of Parkinson’s disease patients based on spectrogram using local binary pattern descriptors”, w: *Journal of Physics: Conference Series* 2153.1, 2022, s. 012014, DOI: [10.1088/1742-6596/2153/1/012014](https://doi.org/10.1088/1742-6596/2153/1/012014)
- [28] Mehmet Bilal Er, Esme Isik i Ibrahim Isik: „Parkinson’s detection based on combined CNN and LSTM using enhanced speech signals with Variational mode decomposition”, w: *Biomedical Signal Processing and Control* 70, 2021, s. 103006, ISSN: 1746-8094, DOI: <https://doi.org/10.1016/j.bspc.2021.103006>
- [29] Zhi-Jing Xu i in.: „Parkinson’s Disease Detection Based on Spectrogram-Deep Convolutional Generative Adversarial Network Sample Augmentation”, w: *IEEE Access* 8, 2020, s. 206888–206900, DOI: [10.1109/ACCESS.2020.3037775](https://doi.org/10.1109/ACCESS.2020.3037775)
- [30] Marek Wodzinski i in.: „Deep Learning Approach to Parkinson’s Disease Detection Using Voice Recordings and Convolutional Neural Network Dedicated to Image Classification”, w: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, s. 717–720, DOI: [10.1109/EMBC.2019.8856972](https://doi.org/10.1109/EMBC.2019.8856972)
- [31] Onur Karaman i in.: „Robust automated Parkinson disease detection based on voice signals with transfer learning”, w: *Expert Systems with Applications* 178, 2021, s. 115013, ISSN: 0957-4174, DOI: <https://doi.org/10.1016/j.eswa.2021.115013>
- [32] Sungho Shin i in.: „Self-Supervised Transfer Learning from Natural Images for Sound Classification”, w: *Applied Sciences* 11.7, 2021, ISSN: 2076-3417, DOI: [10.3390/app11073043](https://doi.org/10.3390/app11073043)
- [33] Paul Faragó i in.: „CNN-Based Identification of Parkinson’s Disease from Continuous Speech in Noisy Environments”, w: *Bioengineering* 10.5, 2023, ISSN: 2306-5354, DOI: [10.3390/bioengineering10050531](https://doi.org/10.3390/bioengineering10050531)

- [34] Alex S. Ozbolt i in.: „Things to Consider When Automatically Detecting Parkinson’s Disease Using the Phonation of Sustained Vowels: Analysis of Methodological Issues”, w: *Applied Sciences* 12.3, 2022, ISSN: 2076-3417, DOI: [10.3390/app12030991](https://doi.org/10.3390/app12030991)
- [35] Daria Hemmerling: „Wykorzystanie sygnału mowy jako źródła informacji diagnostycznej, kontrolnej i prognostycznej w wybranych problemach medycznych związanych z otolaryngologią”, prac. dokt., Akademia Górnictwo-Hutnicza im. Stanisława Staszica w Krakowie, 2018
- [36] Giovanni Dimauro i Francesco Girardi: *Italian Parkinson’s Voice and Speech*, 2019, DOI: [10.21227/aw6b-tg17](https://doi.org/10.21227/aw6b-tg17)
- [37] Juan Rafael Orozco i in.: „New Spanish speech corpus database for the analysis of people suffering from Parkinson’s disease”, w: maj 2014
- [38] Máté Hireš i in.: „Voice-Specific Augmentations for Parkinson’s Disease Detection Using Deep Convolutional Neural Network”, w: *2022 IEEE 20th Jubilee World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, 2022, s. 000213–000218, DOI: [10.1109/SAMI54271.2022.9780856](https://doi.org/10.1109/SAMI54271.2022.9780856)