

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: *inż. 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 rozpośrednia 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.

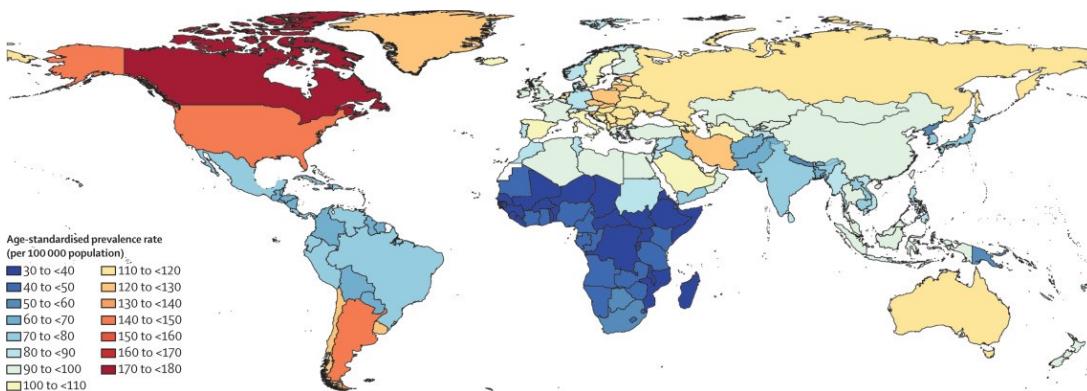
Spis treści

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. Metody diagnozy choroby Parkinsona	19
3.2. Klasyfikacja choroby Parkinsona	21
3.3. Wyzwania związane z systemami automatycznej diagnostyki.....	25
4. Materiał i metoda badawcza	29
4.1. Materiał badawczy.....	29
4.2. Parametryzacja sygnału akustycznego	34
4.3. Metody klasyfikacji	37
4.4. Metody ewaluacji wyników.....	41
5. Wyniki badań	45
5.1. Samogłoska /a/.....	45
5.2. Samogłoska /e/.....	46
5.3. Samogłoska /i/	47
5.4. Samogłoska /o/	49
5.5. Samogłoska /u/	50
5.6. Połączenie samogłosek /a/, /e/, /i/, /o/, /u/	51
5.7. Zbiorcze podsumowanie wyników	52
6. Analiza i interpretacja wyników.....	53
6.1. Samogłoski a informacja diagnostyczna	53
6.2. Architektury CNN a skuteczność klasyfikacji.....	54

6.3. Kierunki rozwoju.....	55
7. Podsumowanie.....	59
Bibliografia	61

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.

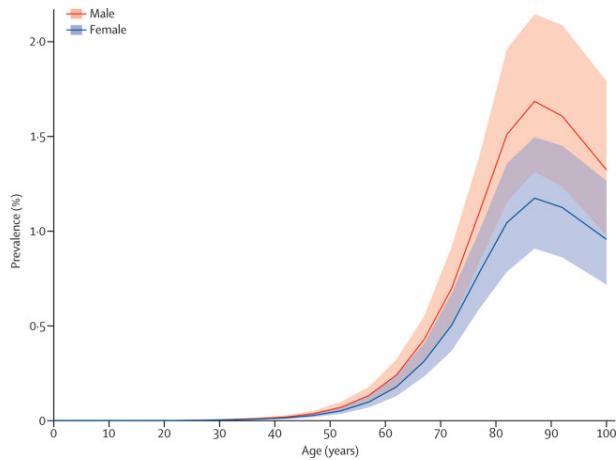


Rys. 1.1. Choroba Parkinsona na świecie [1]

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 tą 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).



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

Każdy może być narażony na ryzyko rozwoju choroby Parkinsona, ale częściej występuje ona u mężczyzn niż u kobiet (Rys.1.2). Statystyki pokazują, że ryzyko zachorowania rośnie wraz z wiekiem, chociaż może ona 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.

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, to w większości przypadków nie występuje ona rodzinna [3].

Proces diagnozowania choroby jest niezwykle złożony i czasochłonny. Nie istnieje obecnie kompletne badanie pozwalające na postawienie pewnej 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 ramach niniejszej pracy przeanalizowano aktualny stan rzeczy oraz zbadano potencjał jednego z proponowanych w literaturze zrozwiązań, dotyczącego 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 diagnostycznych.

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. Na ten moment nie wiadomo co powoduje śmierć neuronów. Zanikają również zakończenia nerwowe, które wytwarzają norepinefrynę, główny przekaźnik chemiczny współczulnego układu nerwowego, który kontroluje 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 objawy są subtelne i kształtują się stopniowo. Często zaczynają się od jednej strony ciała lub nawet jednej kończyny. Niektórzy pacjenci doświadczają pewnych związków 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].

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 i powodowane są przede wszystkim przez deficyt 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 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* (krótkookresowe odchylenie od ustalonych, okresowych charakterystyk sygnału), niższy stosunek harmonicznych do szumów (H/N), mniejszą zmienność częstotliwości i intensywności mowy, 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. 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 dotkniętymi 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). Dostępne są również zaawansowane techniki obrazowania, takie jak PET (pozytonowa emisyjna tomografia) oraz SPECT (tomografia emisyjna pojedynczych fotonów), które pozwalają na obserwację metabolizmu w układzie pozapiramidowym. Międzynarodowe kryteria rozpoznania choroby Parkinsona nie nakładają obowiązku wykonywania badań obrazowych w celu potwierdzenia diagnozy.

c) Test z lewodopą

Test polega na podaniu pacjentowi preparatu z lewodopą. Jeśli następuje poprawa po zażyciu, 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. Nie obserwuje się tych zaburzeń w przypadku zaniku wieloukładowego i postępującego 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 postępujące 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 potwierzonej 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 pozwoli na szybsze i trafniejsze udzielanie pomocy pacjentom, a wcześniejsza diagnoza może przynieść oszczędności finansowe. 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 w pełni skutecznej 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, 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 chorobie 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 natężeniu 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.

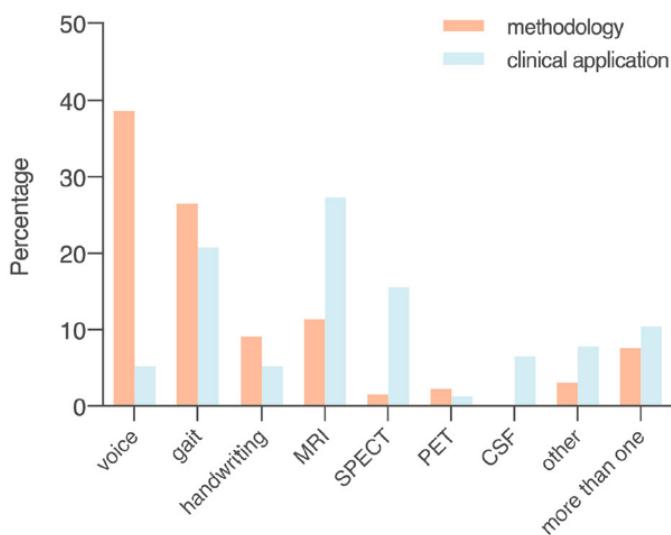
3. Analiza rozwiązań do automatycznej diagnostyki choroby Parkinsona

Diagnoza PD jest powszechnie oparta na obserwacjach lekarskich i ocenie objawów klinicznych, w tym charakterystycznych 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, stan techniki został poszerzony o metody uczenia maszynowego do klasyfikacji PD i osób zdrowych lub pacjentów z podobnymi objawami klinicznymi (np. zaburzeniami ruchu lub innymi zespołami parkinsonowskimi), a rozwój w tej dziedzinie jeszcze się nie zakończył.

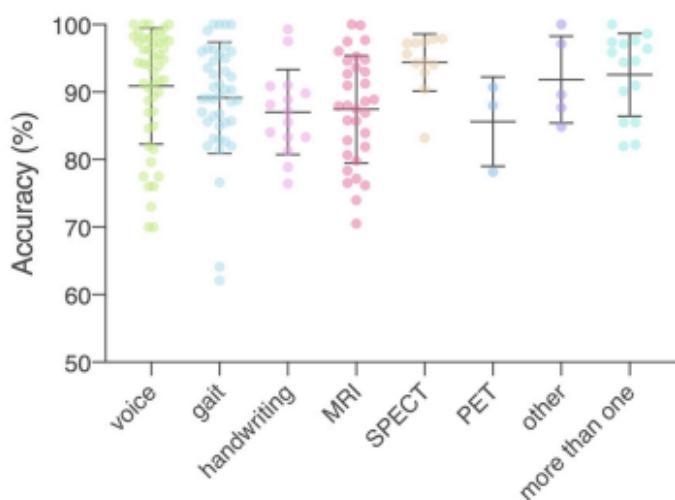
3.1. Metody diagnozy choroby Parkinsona

Diagnozowanie choroby Parkinsona stanowi zadanie złożone ze względu na różnorodność objawów, które dotyczą 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, że zmiany dotyczą przede wszystkim aktywności mózgu. 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. Metody diagnostyki choroby Parkinsona oraz ich częstotliwość wykorzystania [%] [13]

Rys. 3.1 ilustruje zastosowanie wymienionych metod diagnostyki 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 duża. Przyczyny tego zjawiska zostaną dokładniej rozważone w dalszej części pracy. Co ciekawe, jak przedstawiono na Rys. 3.2, detekcja choroby Parkinsona sięga niemalże 100% skuteczności na podstawie badań literaturowych.



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

Podstawą rozwiązań opartych na analizie głosu 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 (syllab),
- 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 nawiązujących 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 nasilony 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. Klasyfikacja choroby Parkinsona

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.

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 tamtego czasu ukazało się kilka dodatkowych publikacji, które pogłębiły ten temat. Na przykład w publikacji [24] zaprezentowano podejście do detekcji choroby Parkinsona, wykorzystując samogłoski o długotrwałej artykulacji oraz architekturę ResNet, pierwotnie dedykowaną do klasyfikacji obrazów. Obliczono spektrogramy nagrań dźwiękowych i wykorzystano je jako dane wejściowe do architektury ResNet wytrenowanej wcześniej na bazach danych ImageNet i SVD. Osiągnięta dokładność przekraczała 90% na zestawie walidacyjnym (PC-GITA, czyli baza danych zawierająca nagrania samogłosek od osób posługujących się kolumbijską odmianą języka hiszpańskiego). Wyniki pokazały, że cechy naucone na naturalnych obrazach potrafią przenieść wiedzę na obrazy reprezentujące spektrogram sygnału głosowego. Co więcej, pokazano, że możliwa jest skuteczna detekcja choroby Parkinsona, wykorzystując tylko cechy oparte na częstotliwościach.

Podejście to zostało zbadane również w artykule [25], gdzie porównano różne architektury CNN do automatycznej identyfikacji PD na podstawie mel-spektrogramów opierając się na uczeniu transferowym. Porównano architektury SqueezeNet1_1, ResNet101 i DenseNet161. Wyniki wykazały, że zaproponowany model oparty na uczeniu transferowym z podejściem fine-tuning zapewnia akceptowną detekcję PD z dokładnością wynoszącą 89,75%.

Później, w 2022 roku autorzy artykułu [26] 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. Mimo niewielkich różnic pomiędzy różnymi samogłoskami najlepszą skuteczność osiągnięto przy uwzględnieniu samogłoski /a/ uzyskując 99% dokładności. 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 [27]. 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 klasyfikatora Random Forest. 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 wagi 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 [28] 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 [29] 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 [30] 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 służyły do wyciągania cech z obrazów mel-spektrogramu sygnałów mowy, a sieć LSTM była 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 nagrani, 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 [31] 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%.

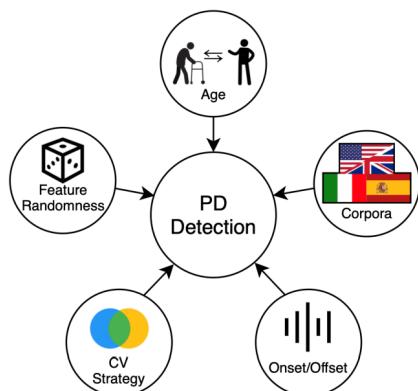
Wszystkie te badania prowadzą do wniosku, że wykorzystanie uczenia maszynowego oraz głębokich sieci neuronowych może istotnie poprawić precyzję i skuteczność diagnostyki choroby Parkinsona na podstawie analizy cech głosowych. Te nowoczesne techniki otwierają nowe horyzonty w kwestii doskonalenia opieki 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.

Wartością dodaną tych badań jest możliwość zastosowania analizy głosowej w procesie diagnozowania choroby Parkinsona, co może umożliwić wcześnie 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.

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 [32] przeprowadzili analizę, wpływu niektórych czynników na wyniki klasyfikacji. Głównym celem było 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 eksperimentalnym 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 [32]

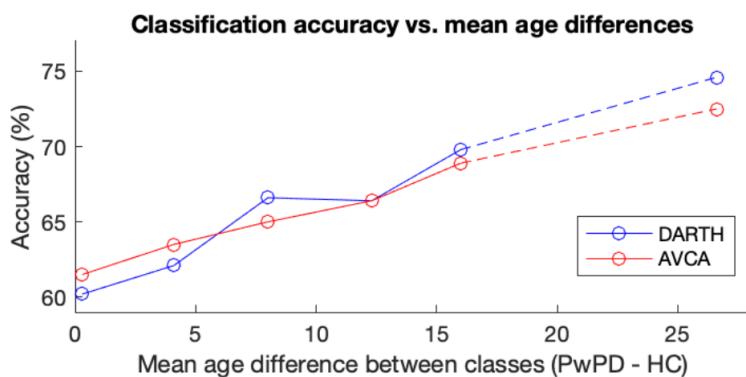
- Pominięcie aspektu tożsamości mówcy przy konstruowaniu zbiorów treningowych i testowych

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 [32] 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 [32]

c) Wpływ losowości cech na dokładność klasyfikacji

W publikacji [32] 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 wydzielenie osobnego 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 [32]. 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 negatywnie na wydajność klasyfikacji.

- e) Wpływ początku i końca nagrani samogłosek na wyniki klasyfikacji

Główna różnicą między korpusami wykorzystywany 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 [32]. 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 [32] 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 nagrań, 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 największe.

Identyfikacja i świadomość wpływu powyższych czynników pozwala na dostosowanie 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 mogłyby zostać wykorzystane w rzeczywistym środowisku.

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 danych 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.
- 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 diagnozy 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 [32].

4. Materiał i metoda badawcza

Przegląd literatury ukazuje obfitość raportów dotyczących identyfikacji choroby Parkinsona na podstawie krótkich fragmentów mowy, takich jak samogłoski, sylaby, czy krótkie słowa i zdania. W tych badaniach wykorzystywane są różnorodne zadania wokalne, metody ekstrakcji cech oraz klasyfikacji. Jednak wiele czynników wpływa na trudność obiektywnego porównania proponowanych rozwiązań, w tym różnice w wykorzystywanych bazach danych.

W niniejszej pracy przeprowadzono porównanie skuteczności wybranych architektur sieci konwolucyjnych oraz ocenić ich przydatność w diagnozowaniu choroby Parkinsona na podstawie różnych samogłosek. Zastosowano połączenie trzech różnych baz danych w celu poszerzenia zbioru o dodatkowe wzorce.

W tym rozdziale przedstawiono szczegółowy opis danych użytych do porównania efektywności i wydajności różnych algorytmów w diagnozowaniu choroby Parkinsona na podstawie mowy. Omówiono proces przygotowania sygnałów oraz parametryzację sygnału akustycznego. Następnie przedstawiono opis algorytmów, które zostały wykorzystane do realizacji postawionych zadań. Schemat przyjętego podejścia został zilustrowany na Rys. 4.1.



Rys. 4.1. Schemat przyjętego podejścia

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 uzyte

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 [32] 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.

W pracy wykorzystano dane, zarejestrowane dla 3 różnych języków: polskiego, włoskiego oraz hiszpańskiego. Wykorzystanie baz danych uwzględniając różne grupy językowe pozwala na analizę uniwersalności narzędzia służącego do automatycznej diagnozy choroby Parkinsona, tak by nie było ono ograniczone tylko do jednej grupy narodowościowej czy geograficznej. Takie podejście pozwala na utworzenie modelu uniwersalnego, który mógłby być używany 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 w ramach realizacji rozprawy doktorskiej we współpracy z Krakowskim Szpitalem Specjalistycznym im. Jana Pawła II [33]. W bazie danych znajdują się 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 podejrzewaną 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 liczebność 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* [34]. 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 oczyszczaniu 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) [35] zawiera nagrania mowy 50 pacjentów z PD i 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

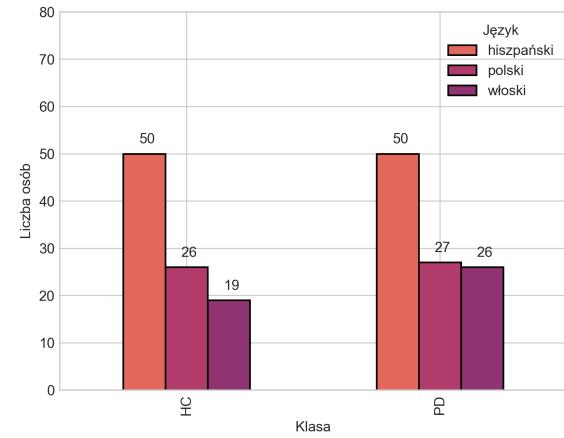
4.1.4. Podsumowanie wykorzystanych baz danych

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.

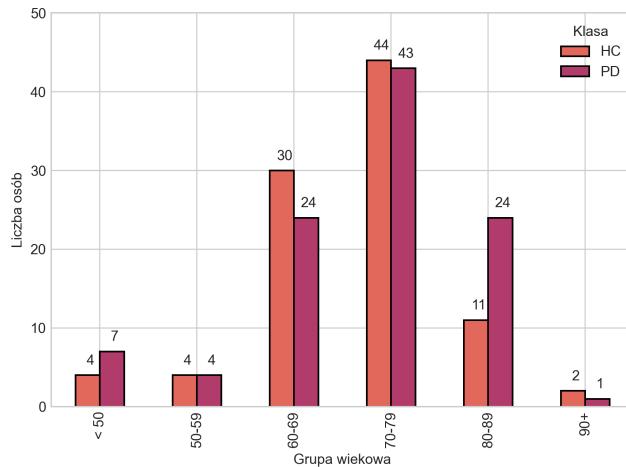
Trudno nie zauważyc znaczącej przewagi liczby nagrań w bazie hiszpańskojęzycznej (Rys. 4.2a). Jednak zrezygnowano z wyrównywania liczby nagrań w poszczególnych językach, z uwagi na to, że baza danych jest już stosunkowo niewielka, i dodatkowe ograniczenie liczby nagrań mogłoby znaczco zmniejszyć jej 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

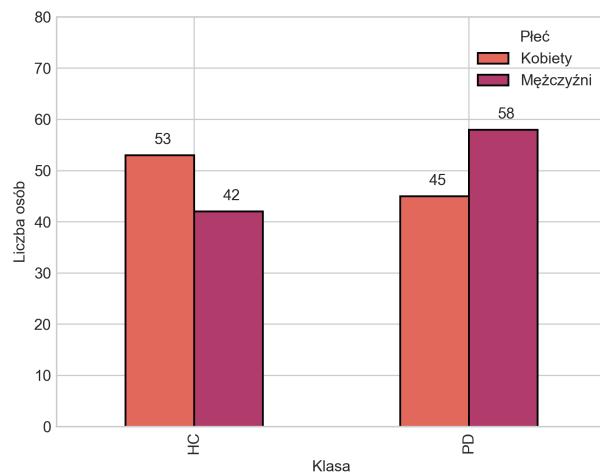
Kategoria	Osoby zdrowe (HC)	Osoby chore (PD)	Razem
Liczba osób	95	103	198
Liczba nagrań	255	229	484
Średnia wieku	$62,18 \pm 8,70$	$62,84 \pm 9,45$	$62,32 \pm 8,99$
Liczba kobiet	53	45	98
Liczba mężczyzn	42	58	100



(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 danych. Proces ten jest niezwykle istotny i 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 wstępniego danych dźwiękowych przed ich podaniem modelowi uczenia maszynowego. W ramach tego procesu usuwane są zakłócenia, szумy i niepożądane. W przypadku 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

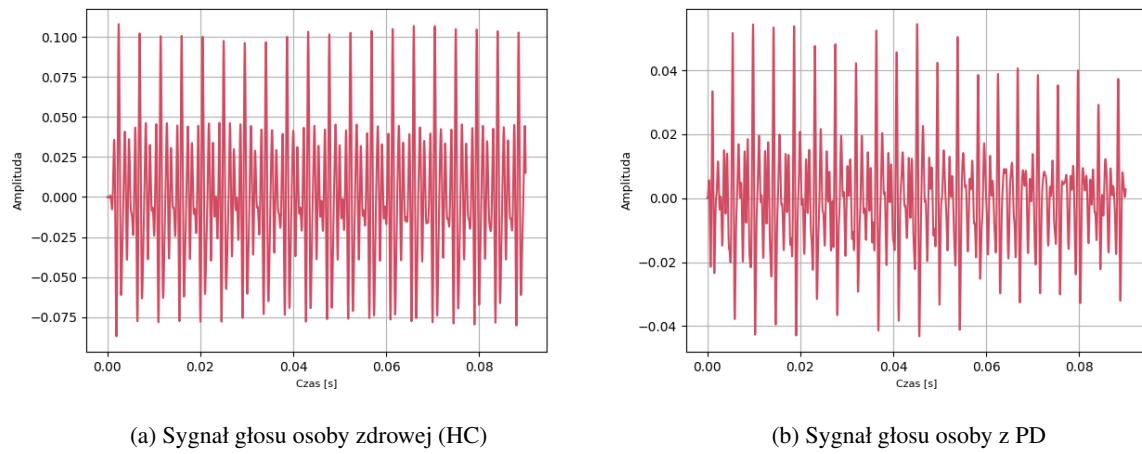
Nagrania zostały przyjęte, tak by nie zawierały początkowych i końcowych fragmentów ciszy. Ten proces eliminuje niepotrzebne fragmenty i skupia analizę jedynie na istotnych fragmentach mowy, co może poprawić skuteczność modelu. W celu automatycznego usunięcia niepożądanych fragmentów skorzystano z pakietu *Librosa* do, 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 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.

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



(a) Sygnał głosu osoby zdrowej (HC)

(b) Sygnał głosu osoby z PD

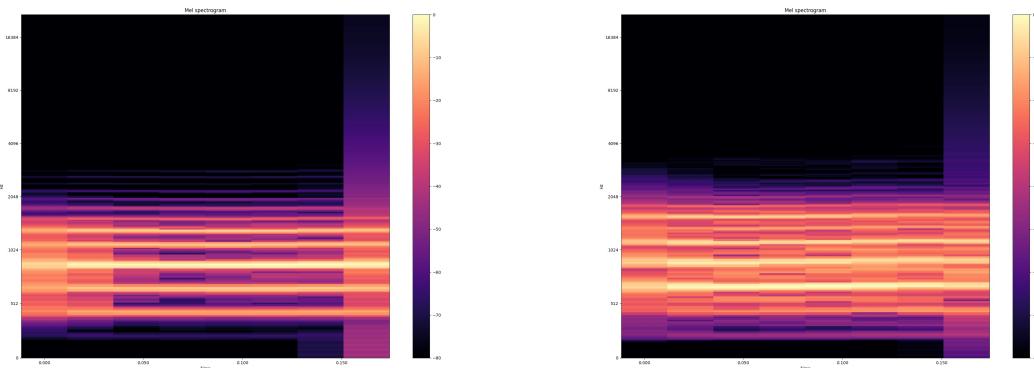
Rys. 4.3. Porównanie przykładowego nagrania osoby zdrowej (HC) oraz chorej (PD), po preprocessingu (hiszpańskojęzyczny zbiór danych, samogłoska /a/)

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 percepcji ludzkiego słuchu. Stosuje się ją, aby lepiej odzwierciedlić sposób, w jaki odbieramy i rozpoznajemy różnice między różnymi tonami dźwięków. Przez uwzględnienie charakterystyk percepcji słuchowej człowieka, mel-spektrogramy dostarczają bardziej reprezentatywnych danych akustycznych dla analizy i klasyfikacji dźwięku. Proces obliczeń mel-spektrogramów został zrealizowany przy użyciu biblioteki języka Python, Librosa. Wygenerowane mel-spektrogramy zostały przedstawione jako obrazy w skali logarytmicznej (Rys. 4.4).

Ważnym aspektem przy wyznaczaniu mel-spektrogramów są parametry 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 reprezentacji dźwięku. Wybór tych wartości został przeprowadzony na podstawie dostępnej literatury oraz własne eksperymenty.

Rozmiar okna czasowego jest jednym z kluczowych parametrów obliczeń melspektrogramów. W kontekście diagnozowania choroby Parkinsona istnieje potrzeba uwzględnienia zarówno krótkotrwałych,



(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 (hiszpanskojęzyczny zbiór danych, samogłoska /a/)

dynamicznych zmian w mowie, jak i dłuższych charakterystyk. Wybór długości okna (2048) został dokonany w celu uzyskania odpowiedniej rozdzielczości w dziedzinie czasu i częstotliwości, co pozwala na dokładniejsze odwzorowanie tych cech w mel-spektrogramie. 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. 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).

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

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 badaniu opisany w publikacji [36] 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 pokazały, że wykorzystane techniki augmentacji istotnie poprawiały 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 [24].

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*). Dla zapewnienia, 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].
- d) Przyspieszenie (ang. *speed-up*). Tak samo, jak spowolnienie dźwięku, losowe przyspieszenie ma na celu zapobieganie dostosowywaniu modelu do prędkości mowy mówcy. Współczynnik przyspieszenia jest losowo wybierany z przedziału [1,2; 2,5].

Augmentacja została przeprowadzona tylko dla zbioru treningowego i walidacyjnego.

4.3. Metody klasyfikacji

Zdecydowano się na graficzne reprezentacje sygnałów, co wymaga też doboru odpowiednich metod klasyfikacji. Popularnym rozwiązaniem 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.

Główną cechą CNN są warstwy konwolucyjne. Wykorzystują one 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. Pomaga to 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. 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 pretrenowanych modeli na dużych zbiorach danych, a następnie dostosowanie ich do konkretnej dziedziny lub zadania.

Konwolucyjne sieci neuronowe 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*. Umożliwia ona wykorzystanie wstępnie wytrenowanych modeli, które zostały nauczane na dużym zbiorze danych, do zadań diagnostycznych z wykorzystaniem ograniczonej ilości dostępnych danych głosowych.

Kluczowym krokiem było wybranie modeli, które miały dostępne wagi 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. Takie podejście 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.

- 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.
- 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.
- d) InceptionV3. Charakteryzuje się wykorzystaniem tzw. *inception modules*, które pozwalają na efektywne wykrywanie wielu skal i rodzajów cech w obrazach. 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. Porównanie różnych klasyfikatorów pozwoli na ocenę ich wydajności i skuteczności w diagnostyce na podstawie głosu.

Po początkowym przeszkoleniu na zbiorze danych ImageNet przeprowadzono proces fine-tuningu na wcześniej przygotowanym zestawie danych. Modele były dostosowywane w celu nauki rozpoznawania specyficznych cech akustycznych charakterystycznych dla pacjentów z chorobą Parkinsona. Proces ten obejmował dodawanie warstw gęstych oraz korzystanie z techniki dropout. Dzięki tym modyfikacjom klasyfikatory stały się bardziej odpowiednie do analizy dźwięku oraz diagnozowania 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 eksperymetach 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 eksperymentalnej 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 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.

Modele były trenowane przez 50 epok, ponieważ zbyt duża liczba epok mogłaby prowadzić do przetrenowania na tak małej ilości danych, co wpłynęłoby negatywnie na ich zdolność do generalizacji na nowe dane. Dodatkowo zastosowano monitorowanie funkcji straty. Ten mechanizm umożliwia wcześnie zatrzymanie treningu, jeśli funkcja straty przestaje się poprawiać przez określoną liczbę epok (w tym przypadku 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.

Inną często wykorzystywaną 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 walidacyjny, 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 (biorąc pod uwagę osoby, nie licząc nagrani). Ważnym aspektem jest fakt, że podział nagrani dla różnych samogłosek przeprowadzono biorąc pod uwagę tożsamość mówcy (podejście *subject-wise*). To znaczy, że na przykład nagrania osoby, która znajduje się w zbiorze treningowym nie mogą wystąpić ani w testowym, ani walidacyjnym. 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. Ostatecznie w zbiorze testowym znalazło się 90 nagrani od 30 osób. W zbiorach walidacyjnym i treningowym liczby te nieco różniły się w różnych iteracjach krosswalidacji, ponieważ nie dla każdego mówcy dysponowano taką samą ilością nagrani. Zbiór walidacyjny zawierał około 45 nagrani, a treningowy około 350.

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 jest konieczność zastosowania odpowiednich technik ewaluacji 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 diagnostowania choroby Parkinsona i polepszania dokładności klasyfikacji, co jest kluczowe dla postępu w dziedzinie medycznej 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) Błędy I i II rodzaju

Jest to popularne narzędzie służące do analizy wyników klasyfikacji. Błąd I rodzaju (ang. *False Positive*) oznacza błędную klasyfikację przypadku, który jest negatywny, jako pozytywny. W kontekście Parkinsona oznacza to, że model zaklasyfikował zdrową osobę jako chorą na chorobę Parkinsona. Błąd II rodzaju (ang. *False Negative*) oznacza błędную klasyfikację przypadku, który jest pozytywny, jako negatywny. W kontekście Parkinsona oznacza to, że model zaklasyfikował osobę chorą na chorobę Parkinsona jako zdrową.

Analiza błędów I i II rodzaju 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 błędego zdiagnozowania osoby zdrowej jako cierpiącej na chorobę Parkinsona. Ważne jest, aby oceniać i minimalizować te błędy w celu uzyskania jak najdokładniejszej klasyfikacji.

W systemach diagnostycznych większe znaczenie ma minimalizacja błędów II rodzaju. W przypadku wykorzystania takiego narzędzia jako wspomagającego proces diagnozy zaklasyfikowanie osoby zdrowej jako chorej spowoduje jedynie dalszą diagnostykę. Natomiast jeśli osoba chora rozpoznana zostanie jako zdrowa, proces jej leczenia rozpocznie się później, co może przynieść różne skutki.

b) 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

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

c) 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

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

d) Czułość (ang. *recall*)

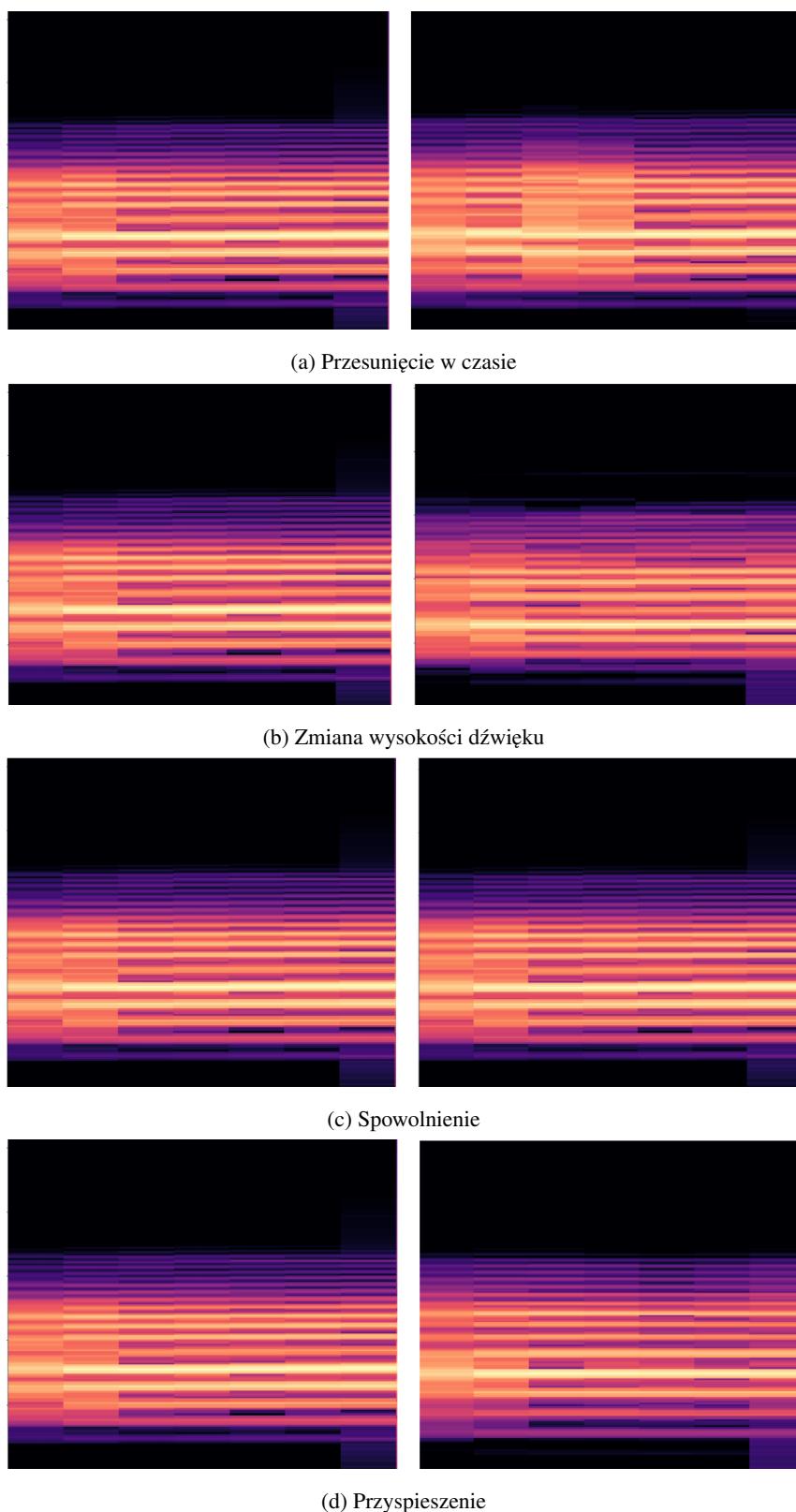
Jest to stosunek prawdziwie pozytywnych predykcji do sumy prawdziwie pozytywnych i fałszywie negatywnych predykcji. Wyrażana jest wzorem 4.4 i określa zdolność modelu do wykrywania wszystkich prawdziwie pozytywnych przypadków.

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

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

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

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



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

5. Wyniki badań

W niniejszym rozdziale zostaną przedstawione rezultaty eksperymentów dotyczących wykorzystania samogłosek /a/, /e/, /i/, /o/ oraz /u/ w procesie automatycznej diagnostyki choroby Parkinsona. Wcześniej omówione zostały teoretyczne podstawy oraz użyte modele i metody badawcze. Niniejsza część pracy zawiera wyniki uzyskane po przeprowadzeniu badań obejmujących porównanie różnych modeli opartych na popularnych architekturach sieci neuronowych, mające na celu ocenę ich skuteczności w identyfikacji choroby Parkinsona. Należy zaznaczyć, że w pracy przedstawiono jedynie najlepsze z uzyskanych wyników. Tabele 5.1, 5.2, 5.3, 5.4 oraz 5.5 przedstawiają średnie z wyników otrzymanych z wykorzystaniem 10-krotnej walidacji krzyżowej.

5.1. Samogłoska /a/

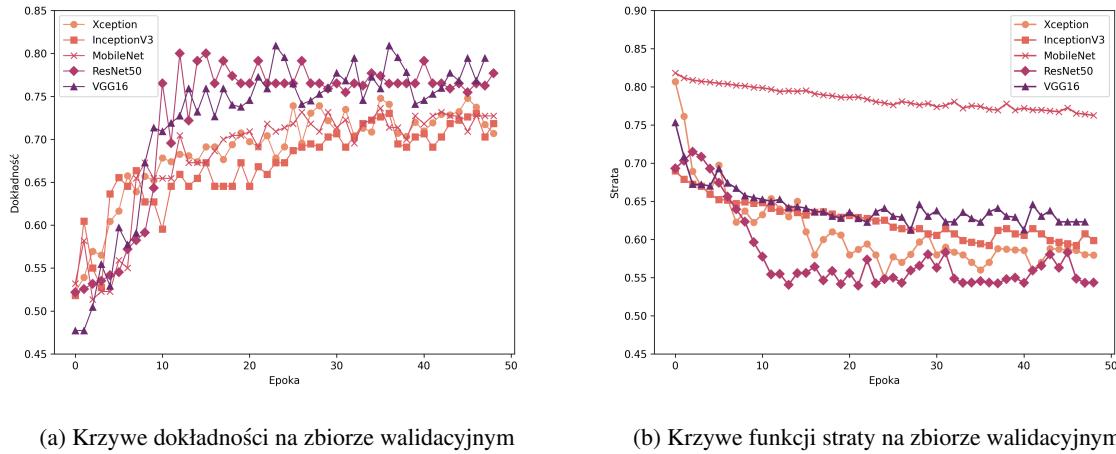
Samogłoska /a/ jest wybierana najczęściej jako podstawa analizy w podobnych procesach badawczych. Ponadto często przy jej wykorzystaniu osiągane są najwyższe wyniki. Podczas eksperymentów w przypadku samogłoski /a/ najlepszą skuteczność uzyskano wykorzystując architekturę VGG16. Każda z badanych metryk (dokładność, czułość, precyzja oraz parametr F1) oscylowała w okolicy 75%. Kolejne dwie architektury, dla których osiągnięto zadowalające wyniki na poziomie ponad 70% to ResNet50 i MobileNetV2. Najgorzej z tym zadaniem, biorąc pod uwagę skuteczność, poradziła sobie architektura Xception.

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

Analiza wartości funkcji straty dla różnych modeli wskazuje na zróżnicowaną skuteczność w procesie minimalizacji błędu. Model ResNet50 osiągnął najniższy wynik straty (0,563), co sugeruje jego lepszą zdolność do generalizacji na dane testowe. Jednakże, modele VGG16 (0,645) i Xception (0,611)

również uzyskały akceptowalne wyniki, podczas gdy model MobileNetV2 (0,750) wypada pod tym względem najsłabiej. Wszystkie z przedstawionych rozwiązań mogą wymagać dalszej optymalizacji lub dostrojenia parametrów w celu poprawy skuteczności. Szczegółowe wartości zostały zawarte w tabeli 5.1.



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

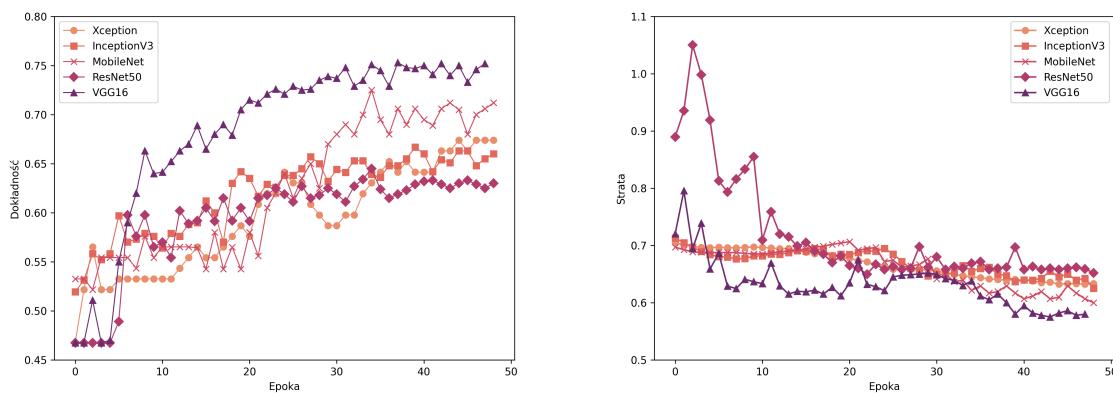
Analizując proces uczenia na zbiorze walidacyjnym (Rys. 5.1), można stwierdzić, że przebiegał on podobnie w większości przypadków. Wartości osiągnięte na tym zbiorze były zbliżone do wyników uzyskanych na zbiorze testowym, co wskazuje na zdolność modeli do generalizacji. Wykresy reprezentujące różne klasyfikatory wykazują podobne tendencje, jednak krzywa funkcji straty w przypadku modelu MobileNetV2 wykazuje pewne odchylenia od pozostałych. To może sugerować, że MobileNetV2 reaguje inaczej na dane walidacyjne, co może mieć wpływ na jego ogólną skuteczność. Na podstawie przedstawionych wyników można wyróżnić model VGG16 jako potencjalnie najlepszy w przypadku diagnostyki choroby Parkinsona na podstawie samogłoski /a/. Dobrym wyborem byłby również niewiele odbiegający ResNet50, wykazujący się niższą wartością funkcji straty.

5.2. Samogłoska /e/

W przypadku samogłoski /e/ otrzymano wyniki nieco gorsze, ale zbliżone do tych uzyskanych w przypadku samogłoski /a/. Tym razem również architektury VGG16 (73%) i MobileNetV2 (70%) osiągnęły najwyższą skuteczność, co wskazuje na ich potencjał w tym zadaniu klasyfikacyjnym. Wartości funkcji straty wahają się od 0,582 (dla VGG16) do 0,659 (dla ResNet50). Jest to istotna informacja, ponieważ niższa wartość straty może sugerować lepszą zdolność modelu do dokładnej klasyfikacji. Warto zwrócić uwagę, że architektura ResNet50 (62%) uzyskała najniższą dokładność w tym przypadku, mimo wysokich wyników dla samogłoski /a/. Raport przedstawiony został w tabeli 5.2.

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

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

Jeśli chodzi o zachowanie modelu w trakcie procesu uczenia, można zauważyc, że było ono podobne we wszystkich przypadkach, co zostało zobrazowane na Rys. 5.2. Modele osiągały względna stabilność w różnych momentach procesu uczenia i różniły się tempem dostosowywania się do danych treningowych oraz początkowymi wartościami funkcji straty w kilku pierwszych epokach. Jednak na końcu procesu uczenia wartości te były zbliżone do siebie i odzwierciedlały te uzyskane na zbiorze testowym. Nie zaobserwowano wyraźnego przeuczenia (ang. *overfittingu*) modeli, co jest pozytywnym wynikiem eksperymentu. Na tle pozostałych ponownie najlepiej sprawdził się model VGG16.

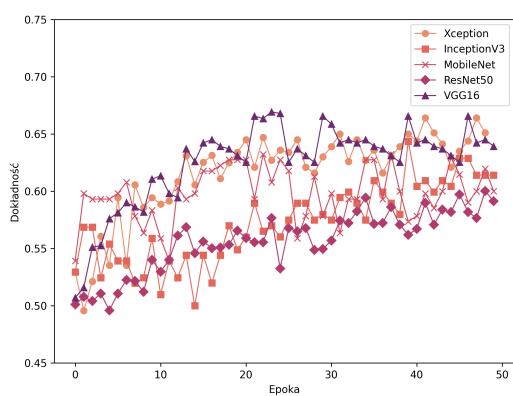
5.3. Samogłoska /i/

W przypadku samogłoski /i/ można zauważyc, że wyniki uległy pogorszeniu w porównaniu do samogłoski /a/ oraz /e/. Najlepsze rezultaty uzyskano przy wykorzystaniu architektur MobileNetV2 i VGG16, gdzie dokładność wyniosła około 63%. Warto zaznaczyć, że jest to znacznie niższy poziom dokładności w porównaniu do wyników uzyskanych dla samogłoski /a/. Szczególnie niski wynik obserwuje się ponownie dla modelu ResNet50, który osiągnął tylko około 59% dokładności.

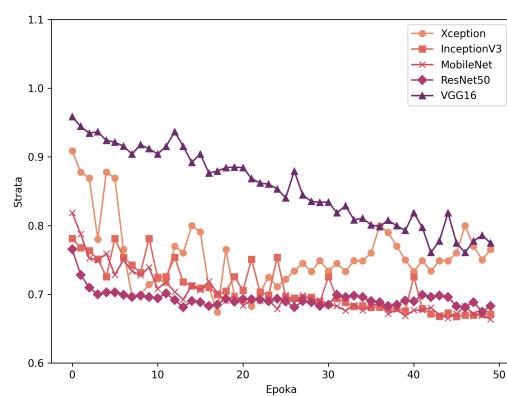
Wartości funkcji straty są natomiast bardzo wysokie, w zakresie od 0,656 dla MobileNetV2 do 0,775 dla VGG16. To może sugerować, że informacja diagnostyczna zawarta w samogłosce /i/ jest trudniej dostępna w porównaniu do innych samogłosek, co wpływa na wyższe wartości straty i niższą dokładność. Możliwe jest, że inne samogłoski mogą dostarczać bardziej wyraźnych i łatwiejszych do interpretacji sygnałów diagnostycznych. Szczegółowe wartości przedstawia tabela 5.3.

Tabela 5.3. Wyniki otrzymywane 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	0.705	0.686	0.656



(a) Krzywe dokładności na zbiorze walidacyjnym



(b) Krzywe funkcji straty na zbiorze walidacyjnym

Rys. 5.3. Porównanie krzywych uczenia dla różnych klasyfikatorów na zbiorze walidacyjnym (samogłoska /i/)

Na Rys. 5.3 przedstawiono proces uczenia na zbiorze walidacyjnym dla samogłoski /i/. Nie zaobserwowano oznak przeuczenia, a wykresy procesu uczenia są zbliżone do siebie. Jedynie wartość funkcji straty jest znacznie wyższa w przypadku modelu VGG16 niż w pozostałych przypadkach. Warto zauważyć, że proces uczenia wydaje się mniej stabilny w przypadku modeli Inception i Xception, co może wpływać na gorsze wyniki.

Mimo że żaden z tych modeli nie jest zalecany do wykorzystania w rzeczywistym środowisku, a wyniki dla innych samogłosek są wyższe, to nadal istnieje możliwość wykorzystania /i/ jako podstawy klasyfikacyjnej. Konieczne jest jednak dokładniejsze zgłębienie tematu i specjalne dostosowanie metody.

5.4. Samogłoska /o/

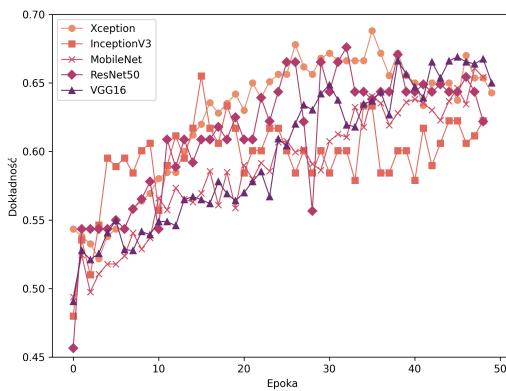
Wyniki uzyskane w przypadku samogłoski /o/ wykazują znaczne podobieństwo do tych uzyskanych w przypadku samogłoski /i/. Dokładność klasyfikacji oscyluje między 61% a 65%, a funkcje straty są zbliżone (około 0,67). Analizując wykresy procesu uczenia na Rys. 5.4, również można zaobserwować podobne zjawisko.

Na podstawie tych wyników nie da się wyróżnić optymalnego rozwiązania. Konieczne jest pogłębianie analizy o więcej danych lub być może inne metody ekstrakcji cech. Z pewnością istnieją markery, które pozwalają na odróżnienie osób chorych na PD od osób zdrowych na podstawie jednostki /i/, jednak nie są one tak wyraźne przy zastosowaniu wykorzystywanych w tym badaniu metod.

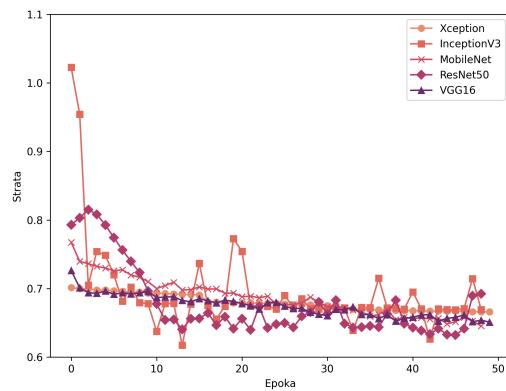
Podobieństwo wyników między samogłoskami /o/ i /i/ może sugerować, że te dwie samogłoski zawierają podobne cechy lub informacje diagnostyczne dla choroby Parkinsona. Jednak warto zwrócić uwagę, że choć wyniki są zbliżone, to nie są one identyczne. Istnieje subtelna różnica w dokładności klasyfikacji oraz wartościach funkcji straty, co może wynikać z różnic w akustyce lub wymowie tych dźwięków.

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



(a) Krzywe dokładności na zbiorze walidacyjnym



(b) Krzywe funkcji straty na zbiorze walidacyjnym

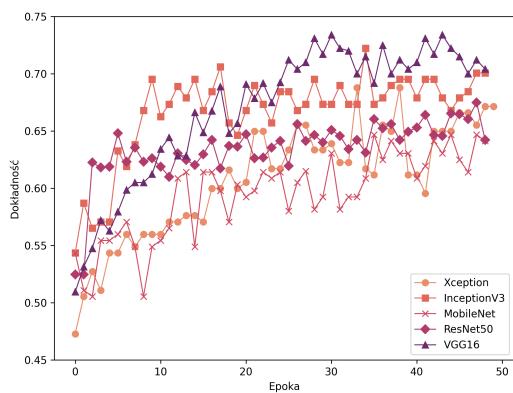
Rys. 5.4. Porównanie krzywych uczenia dla różnych klasyfikatorów na zbiorze walidacyjnym (samogłoska /o/)

5.5. Samogłoska /u/

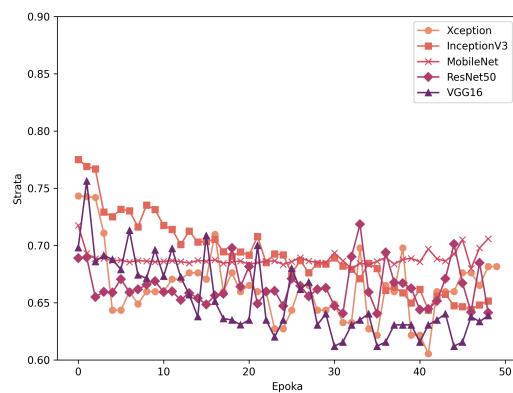
W przypadku samogłoski /u/ obserwacje sugerują podobieństwo wyników do tych uzyskanych dla samogłoski /e/. Tutaj również model VGG16 osiąga najlepsze wyniki, osiągając poziom skuteczności na poziomie 72%. Natomiast wyniki pozostałych modeli wahają się w zakresie od 64% do 70%. Należy zwrócić uwagę na istotną różnicę w funkcji straty pomiędzy MobileNetV2 a pozostałymi modelami. Wykazuje on znacznie wyższą wartość funkcji straty (0,750) w porównaniu do pozostałych modeli, które osiągają wartości w zakresie od 0,637 do 0,685. Ta różnica może być istotna w kontekście oceny wydajności modeli. Szczegółowe wartości przedstawia tabela 5.5.

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



(b) Krzywe funkcji straty na zbiorze walidacyjnym

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

Analizując procesy uczenia testowanych modeli (Rys. 5.5), można zauważyc znaczne podobieństwo. Warto zaznaczyć, że w przypadku MobileNetV2 funkcja straty pozostaje praktycznie stała przez cały proces uczenia, co nie jest pożądany zjawiskiem.

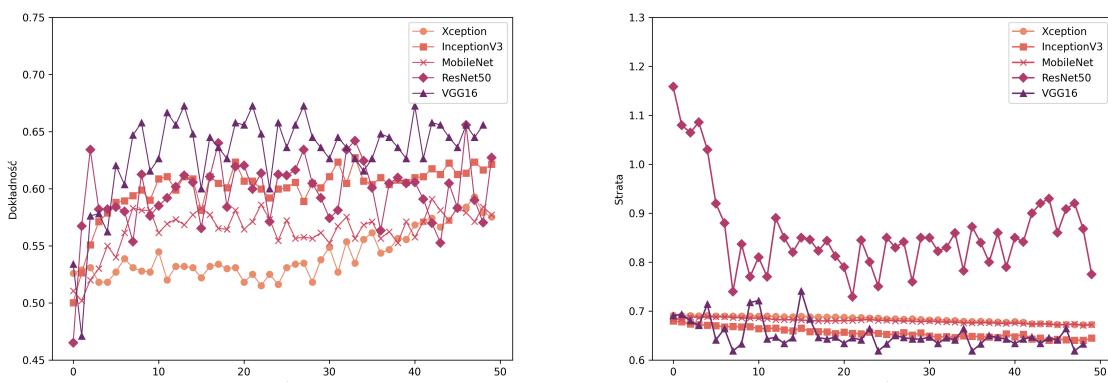
5.6. Połączenie samogłosek /a/, /e/, /i/, /o/, /u/

Zdecydowano się również zbadać skuteczność modeli trenowanych na zbiorze, który został stworzony przez połączenie wszystkich dotychczas badanych samogłosek. Teoretycznie, taka praktyka mogła przyczynić się do poprawy wyników poprzez zwiększenie różnorodności zbioru danych. Jednak nie przyniosła ona oczekiwanych efektów. Może to wynikać z faktu, że rozszerzenie bazy nie zmieniło zestawu osób mówiących, a także możliwe, że markery choroby różnią się w zależności od wypowiadanej samogłoski. Wyniki uzyskane w tym przypadku kształtują się w zakresie od 58% do 63%, a żadna z innych metryk nie wykazuje istotnych odchyleń od tego zakresu. Warto również zwrócić uwagę na wartość funkcji straty, która utrzymuje się na poziomie około 0,7 z wyjątkiem modelu ResNet50, gdzie wartość ta znacznie odbiega i wynosi 0,970. Cały raport został przedstawiony w tabeli 5.6.

Tabela 5.6. Wyniki otrzymane dla połączenia samogłosek /a/, /e/, /i/, /o/ oraz /u/

Model	VGG16	Resnet50	Xception	InceptionV3	MobileNetV2
Accuracy	0.630	0.577	0.626	0.615	0.638
Precision	0.652	0.578	0.627	0.621	0.635
Recall	0.622	0.575	0.625	0.611	0.637
F1-score	0.598	0.567	0.631	0.607	0.639
Loss	0.692	0.970	0.654	0.677	0.645

Znacznie odmienną wartość funkcji straty dla modelu ResNet50 można zobaczyć na Rys. 5.6. Natomiast pozostałe modele wykazują bardzo niewielką zmienność w trakcie procesu uczenia, co przejawia się w niewielkim spadku tej wartości. To zjawisko może sugerować, że ResNet50 doświadczało pewnych trudności lub przeszkód w procesie uczenia, które nie występowały w przypadku innych modeli.



Rys. 5.6. Porównanie krzywych uczenia dla różnych klasyfikatorów na zbiorze walidacyjnym (połączenie samogłosek /a/, /e/, /i/, /o/ oraz /u/)

5.7. Zbiorcze podsumowanie wyników

Na podstawie przedstawionych do tej pory wyników, zebrano najlepsze z nich w Tab. 5.7, aby ułatwić ich porównanie. Najlepsze osiągnięcia zostały zarejestrowane dla samogłosek /a/, /e/ oraz /u/, gdzie dokładność przekroczyła 70%. W przypadku /i/ oraz /o/, a także połączenia wszystkich samogłosek, wyniki były znacznie niższe i utrzymywały się na poziomie około 65%. Pozostałe metryki są zbliżone, co wskazuje na podobną jakość rozpoznawania każdej z klas. Oznacza to, że prawdopodobieństwo popełnienia błędu I jak i II rodzaju są bardzo podobne. Warto zaznaczyć, że w analizowanym problemie badawczym szczególnie istotna jest minimalizacja błędu II rodzaju.

Jeśli chodzi o architektury, to najczęściej najlepiej radził sobie model VGG16, osiągając najwyższą skuteczność w przypadku 4 z 6 analizowanych przypadków. Obiecująco wypadły także modele Xception i MobileNetV2 w przypadku samogłosek /i/ oraz /o/.

Tabela 5.7. Najlepsze z otrzymanych wyników dla poszczególnych wariantów badanych samogłosek

Samogłoska	/a/	/e/	/i/	/o/	/u/	wszystkie samogłoski
Model	VGG16	VGG16	Xception	VGG16	VGG16	MobileNetV2
Accuracy	0.751	0.733	0.641	0.653	0.724	0.638
Precision	0.752	0.729	0.638	0.654	0.730	0.635
Recall	0.751	0.733	0.642	0.649	0.715	0.637
F1-score	0.749	0.75	0.643	0.650	0.712	0.639
Loss	0.645	0.582	0.705	0.674	0.654	0.645

Warto zaznaczyć, że funkcja straty utrzymywała się na stosunkowo wysokim poziomie we wszystkich analizowanych przypadkach, co może sugerować pewne trudności lub wyzwania w procesie uczenia. Istnieje możliwość, że optymalizacja modeli mogłaby być bardziej precyzyjna, aby osiągnąć niższą wartość funkcji straty i poprawić skuteczność klasyfikacji. Dalsza analiza i eksperymenty mogą być potrzebne w celu doskonalenia wydajności modeli.

6. Analiza i interpretacja wyników

Przedstawione wyniki, mimo że nie osiągnęły jeszcze poziomu zadowalającego dla zastosowań w rzeczywistych warunkach, stanowią obiecujący punkt wyjścia. Warto podkreślić, że analiza została przeprowadzona na niewielkiej grupie badawczej, co wprowadza pewne ograniczenia. Zebrano nagrania od blisko 200 uczestników zróżnicowanych pod względem narodowości, w tym 103 osoby z diagnozą choroby Parkinsona. Jednakże, aby pełniej oddać różnorodność i reprezentatywność populacji, konieczne jest zgromadzenie większej liczby próbek.

Podjęcie tego wyzwania jest kluczowe, gdyż badanie głębokich metod uczenia maszynowego w kontekście diagnostyki choroby Parkinsona to złożone zagadnienie badawcze. Analiza literatury naukowej wskazuje, że istnieją możliwości konstrukcji bardziej efektywnych narzędzi, korzystających z zaawansowanych metod uczenia maszynowego.

Mimo to, uzyskany wynik na poziomie 75% jest bardzo obiecujący i otwiera perspektywy wykorzystania głębokich metod uczenia maszynowego w diagnozowaniu choroby Parkinsona. Dalsze badania, rozszerzenie bazy danych oraz ulepszenia w modelach mogą przyczynić się do osiągnięcia jeszcze lepszych rezultatów w przyszłości, co byłoby znaczącym osiągnięciem w dziedzinie medycyny.

Warto podkreślić, że badania opierały się na połączeniu trzech zupełnie różnych baz danych, które reprezentowały różne narodowości. Ten interdyscyplinarny i międzynarodowy charakter podejścia wskazuje na potencjał stworzenia narzędzia diagnostycznego o uniwersalnym zastosowaniu. To otwiera perspektywy na rozwinięcie narzędzia, które będzie skuteczne w diagnozowaniu choroby Parkinsona niezależnie od narodowości pacjentów. Ten aspekt pracy jest niezwykle obiecujący i daje nadzieję na przyszłość również z biznesowego punktu widzenia.

Przeprowadzone badania umożliwiają również wstępna analizę przydatności różnych samogłosek i architektur CNN, co zostanie rozwinięte w dalszej części rozdziału.

6.1. Samogłoski a informacja diagnostyczna

Przeprowadzone badania wykazały, że najlepsze wyniki osiągnięto dla samogłoski /a/ (75%), /e/ (73%) oraz /u/ (72%), a znacznie słabsze dla /i/ (64%) oraz /o/ (65%). Połączenie wszystkich analizowanych samogłosek w jeden zbiór nie przyniosło poprawy wyników, mimo znaczącego zwiększenia zbioru uczącego. Może to wiązać się z różnym sposobem przekazywania przez nie informacji diagnostycznej.

W badaniu [37] przeanalizowano nagrania samogłosek /a/, /e/, /i/ oraz /u/ u 7 pacjentów i 7 pacjentek. Analiza wykazała, że charakterystyka tych nagrań różni się zależnie od płci, ale również od wypowiadanej samogłoski. Biorąc pod uwagę powyższe, rozwiązanie do automatycznej diagnostyki PD powinno skupiać się na jednej lub uwzględniać informację o rodzaju wykorzystanej samogłoski.

Nie spotkano się do tej pory z publikacją, która w bezpośredni i rzetelny sposób porównywałaby potencjał poszczególnych samogłosek w kontekście diagnostyki choroby Parkinsona. Natomiast w badaniu, którego głównym celem była ocena augmentacji danych [36] w tym zagadnieniu, pośrednio oceniono potencjał różnych samogłosek w diagnostyce. Najlepsze wyniki osiągnięto przy wykorzystaniu samogłoski /a/ (92%), a najgorsze w przypadku /i/ (74%). Zależność ta pokrywa się z przeprowadzonymi w tej pracy eksperymentami. Ponadto samogłoska /a/ jest najczęściej wykorzystywana w podobnych badaniach, więc gdyby istniała potrzeba wskazania najlepszej podstawy diagnostycznej spośród samogłosek to wskazano by samogłoskę /a/, mimo że jej przewaga nad /e/ nie jest znacząca według przeprowadzonych badań. Oczywiście konieczne jest przeprowadzenie dalszych badań, z wykorzystaniem znacznie większej bazy danych. Mimo to każda z samogłosek wykazuje potencjał wykorzystania w rozwiązaniu do automatycznej diagnostyki choroby Parkinsona na podstawie głosu.

6.2. Architektury CNN a skuteczność klasyfikacji

W niniejszej pracy porównano potencjał wykorzystania 5 różnych klasyfikatorów pod kątem wykorzystania ich jako narzędzia do diagnostyki choroby Parkinsona na podstawie nagrań głosowych samogłosek. Ze względu na niewielką ilość danych użyto podejścia znanego jako transfer learning, którego główną zaletą jest korzystanie ze wcześniej wytrenowanych wag, zamiast losowych wartości. Takie podejście było już badane w literaturze naukowej i wykazało duży potencjał. Dlatego zdecydowano się wybrać 5 różnych architektur CNN i przeprowadzić ich porównanie. Ocena różnych architektur sieci neuronowych w kontekście detekcji anomalii w sygnale mowy u osób z chorobą Parkinsona stanowi ważny aspekt badawczy. Otrzymane wyniki wskazują na istotne różnice w skuteczności tych architektur, co może dostarczyć cennych wskazówek dla przyszłych badań w tej dziedzinie.

Na podstawie przeprowadzonych badań można stwierdzić, że VGG16 wyróżnił się jako najlepsza architektura, osiągając wysokie wyniki dla 4 z 6 testowanych wariantów samogłosek (/a/, /e/, /o/ oraz /u/). Może to sugerować, że jest ona szczególnie skuteczna w wykrywaniu anomalii w sygnale mowy związanych z chorobą Parkinsona w przypadku samogłosek. VGG16 to głęboka architektura konwolucyjna, która składa się z 16 warstw konwolucyjnych i warstw poolingowych. Ta głębokość pozwala na efektywne wykrywanie i ekstrakcję cech złożonych, co może być istotne w rozpoznawaniu subtelnych zmian w sygnale mowy związanych z chorobą Parkinsona. Ponadto wiele warstw konwolucyjnych jest stosowanych po sobie. To umożliwia wielokrotne przekształcanie sygnału wejściowego, co pozwala na wyodrębnienie bardziej abstrakcyjnych cech. W przypadku sygnału mowy może to pomóc w wykryciu subtelnych zmian w akustycie lub innym aspekcie sygnału związanym z chorobą Parkinsona. VGG16 radzi sobie dobrze w przypadku większości wariantów samogłosek, może to sugerować, że ma zdolność

do generalizacji cech istotnych dla detekcji anomalii w sygnale mowy osób z chorobą Parkinsona. To jest ważne, ponieważ różne warianty samogłosek mogą mieć różne cechy akustyczne.

Xception i MobileNet również wykazały się obiecująco, osiągając dobre wyniki w niektórych przypadkach (odpowiednio /o/ oraz wszystkie samogłoski). Choć nieco mniej skuteczne niż VGG16, nadal mogą być rozważane jako przyzwoite alternatywy. Ich dalsza eksploracja może zaowocować jeszcze lepszymi wynikami.

ResNet okazał się mniej skuteczny w porównaniu do pozostałych architektur, co może sugerować, że ta konkretna architektura może nie być odpowiednia do tego konkretnego zadania. Co ciekawe, obiecujące wyniki wykazał dla samogłoski /a/ i był już z powodzeniem wykorzystywany w innych badaniach dotyczących tego zagadnienia.

Inception wydaje się mieć wyniki średnie, co sugeruje, że może być skuteczna w niektórych przypadkach, ale niekoniecznie jest najlepszą opcją w każdym scenariuszu.

Należy zwrócić uwagę, że efektywność tych architektur może zależeć od konkretnej charakteryzacji sygnału mowy i konkretnego rodzaju anomalii związanych z chorobą Parkinsona. Dlatego też warto kontynuować badania w celu dalszej optymalizacji i dostosowania tych architektur do specyfiki tego zadania.

Warto również wziąć pod uwagę, że dalszy *fine-tuning* tych architektur mógłby jeszcze poprawić skuteczność. Ponadto eksploracja innych architektur sieci neuronowych lub technik uczenia maszynowego może również przynieść korzyści w dalszych badaniach w tej dziedzinie.

Przeprowadzone badania dostarczają istotnych informacji na temat skuteczności różnych architektur sieci neuronowych w detekcji anomalii w sygnale mowy osób z chorobą Parkinsona. Wskazują one na potencjalnie obiecujące wyniki VGG16, Xception i MobileNet, ale sugerują również potrzebę dalszych badań i optymalizacji w tej dziedzinie. Warto jednak zaznaczyć, że wybór architektury zależy od konkretnego zadania i zbioru danych. Wyniki mogą różnić się w zależności od charakterystyki danych i konkretnej metody przetwarzania sygnału mowy. Dlatego eksperymentowanie z różnymi architekturami i optymalizacja modelu dla konkretnego zadania jest ważnym krokiem w badaniach nad detekcją anomalii w sygnale mowy u osób z chorobą Parkinsona.

6.3. Kierunki rozwoju

Diagnostyka choroby Parkinsona na podstawie analizy głosu stanowi fascynujące pole badań, które obiecuje rewolucję w wykrywaniu tej choroby w jej wczesnym stadium. Niniejsza praca nie stanowi rozwiązania, które w zadowalającym stopniu mogłyby służyć jako narzędzie diagnostyczne. Natomiast bada potencjał wykorzystania metod głębokiego uczenia maszynowego w tej dziedzinie. Osiągnięte wyniki wskazują bardziej kierunki dalszej eksploracji zagadnienia niż odpowiedź na stawiane pytania.

Pierwszym istotnym kierunkiem rozwoju jest zwiększenie rozmiaru bazy danych. Przeprowadzone badania bazowały na próbkach od pacjentów z różnych narodowości, jednakże, aby narzędzie to stało się

bardziej uniwersalne, konieczne jest uwzględnienie większej liczby danych z różnych kultur i grup wiekowych. Rozszerzenie bazy danych pozwoli na uzyskanie bardziej reprezentatywnego zbioru, co może istotnie wpływać na zdolność modeli do diagnozowania choroby Parkinsona. Markery choroby Parkinsona są zagadniem bardzo skomplikowanym i nie da się na podstawie 100 pacjentów stworzyć uniwersalnego i skutecznego narzędzia diagnostycznego. Należy również pamiętać, że zmiany w głosie nie tylko dotyczą jedynie 89% osób chorych, ale też objawiają się w różnym nasileniu na różnych etapach choroby. Pogłębiona augmentacja danych może przyczynić się do stworzenia bardziej zróżnicowanego zbioru treningowego, co jest kluczowe w kontekście generalizacji modeli. Rozważenie bardziej zaawansowanych technik augmentacji, takich jak zmienne warunki akustyczne czy emulacja różnych akcentów, może poprawić zdolność modeli do rozpoznawania subtelnych zmian w głosie pacjentów.

Kolejnym krokiem jest eksploracja różnych ustawień spektrogramów i mel-spektrogramów. Parametry takie jak rozdzielcość czasowa i częstotliwościowa mają znaczący wpływ na jakość danych wejściowych dla modeli. Dokładna optymalizacja tych ustawień może pomóc w wydobyciu bardziej istotnych informacji z danych i zwiększyć efektywność diagnozy. Takie eksperymenty zostały wstępnie przeprowadzone na cele tej pracy, jednak konieczne jest ich pogłębienie. W literaturze również nie ma informacji na temat badań, które definiowałyby tzw. *złoty środek* tego zagadnienia. Zarówno jeśli chodzi o wybór między spektrogramami a mel-spektrogramami, ich parametry, jak i sama długość nagrania.

Dla dalszego rozwoju narzędzia diagnostycznego kluczowa jest również eksploracja innych i architektur CNN. Oprócz modeli CNN warto rozważyć wykorzystanie zaawansowanych lub specjalizowanych modeli, a także podejść opartych na uczeniu transferowym, być może trenowanych wstępnie na innych zbiorach danych. Warte rozważenia byłaby dalsza eksploracja podejścia przedstawionego w [24]. Polega ono na wstępny trenowaniu modeli na większej bazie danych dotyczącej nagrań głosowych pacjentów z różnymi schorzeniami (w tym przypadku była to baza SVD), a następnie dotrenowanie modelu na danych dotyczących choroby Parkinsona. Kombinacja różnych architektur również może pomóc w osiągnięciu wyższej dokładności diagnozy. Eksperimentowanie z technikami uczenia zespołowego, które łączą wyniki z różnych modeli, może przynieść korzyści w postaci poprawy ogólnej wydajności i niezawodności diagnostyki. Uczenie zespołowe pozwala na korzystanie z różnych perspektyw i źródeł informacji, co może znacząco zwiększyć skuteczność narzędzia.

Kontynuacja procesu *fine-tuningu* wykorzystanych w pracy modeli może przynieść dalsze doskonalenie i dostosowanie do specyficznych cezur diagnostycznych. Dalszy rozwój w tym obszarze może prowadzić do bardziej precyzyjnych i efektywnych narzędzi diagnostycznych. Dużym problemem podczas prób okazał się overfitting, możliwe, że połączenie rozszerzenia bazy danych wraz z dalszym dostrajaniem modeli przyniosłoby znacznie lepsze efekty.

Ostatnim, ale nie mniej istotnym kierunkiem rozwoju, jest uwzględnienie dodatkowych informacji meta. Dane takie jak płeć, wiek czy język mówiącej osoby mogą stanowić cenne informacje w procesie diagnostycznym. Są to informacje, które wpływają na przykład na sposób wydobywania dźwięku i jego wysokość. Dodanie takich cech mogłoby znacząco ułatwić naukę modelu. Jednakże, aby uniknąć błędów

i utrzymania spójności, konieczne jest uzyskanie takich informacji dla wszystkich próbek w zbiorze danych.

Kierunki rozwoju w diagnostyce choroby Parkinsona na podstawie analizy głosu są obiecujące i otwierają perspektywy na stworzenie narzędzia o globalnym zastosowaniu. Kontynuacja badań w opisanych obszarach może przyczynić się do osiągnięcia jeszcze lepszych wyników i poprawić jakość opieki nad pacjentami cierpiącymi na tę chorobę. Komputerowe wykrywanie choroby Parkinsona umożliwiłoby przesiewowe badania populacji oraz częste monitorowanie, dostarczając bardziej obiektywny pomiar obaw, co przyniesie korzyści zarówno pacjentom, jak i dostawcom opieki zdrowotnej.

7. Podsumowanie

Osiągnięte wyniki znacznie odbiegają od najlepszych rozwiązań przedstawionych w literaturze wykorzystujących podobne metody (ponad 90%). Powodem tej rozbieżności jest prawdopodobnie sposób przygotowania oraz ilość i jakość danych. Większość publikacji pomija aspekt preprocessingu lub opisuje go w niewystarczającym stopniu, dlatego odtworzenie raportowanych skuteczności nie jest możliwe. Należałoby poświęcić temu zagadnieniu jeszcze więcej uwagi, ponieważ wydaje się to być to niezwykle ważnym aspektem tego zagadnienia.

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: „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] 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)
- [25] 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>
- [26] 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>
- [27] 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)
- [28] 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>
- [29] 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)
- [30] 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>
- [31] 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)
- [32] 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)
- [33] Daria Hemmerling: „Wykorzystanie sygnału mowy jako źródła informacji diagnostycznej, kontrolnej i prognostycznej w wybranych problemach medycznych związanych z otolaryngologią”, prac. dokt., Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, 2018

- [34] Giovanni Dimauro i Francesco Girardi: *Italian Parkinson's Voice and Speech*, 2019, DOI: [10.21227/aw6b-tg17](https://doi.org/10.21227/aw6b-tg17)
- [35] Juan Rafael Orozco i in.: „New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease”, w: maj 2014
- [36] 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)
- [37] Young-Im Bang i in.: „Acoustic characteristics of vowel sounds in patients with Parkinson disease”, w: *NeuroRehabilitation* 32, maj 2013, s. 649–54, DOI: [10.3233/NRE-130887](https://doi.org/10.3233/NRE-130887)