



**Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie**

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

**KATEDRA ENERGOELEKTRONIKI I AUTOMATYKI  
SYSTEMÓW PRZETWARZANIA ENERGII**

## **Praca dyplomowa inżynierska**

**Wykorzystanie systemów uczących się do  
detekcji patologii głosu**

**The use of machine learning in voice pathology  
detection**

Autor:

Kierunek studiów:

Opiekun pracy:

Barotosz Tyński

Elektrotechnika

dr inż. Mirosław Socha

Kraków, 20 grudnia 2018

*Upředzony 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 rozpowszechnia bez podania nazwiska lub pseudonimu twórcy cudzy utwór w wersji oryginalnej albo w postaci opracowania, artystyczne wykonanie albo publicznie zniekształca taki utwór, artystyczne wykonanie, fonogram, wideogram lub nadanie.”, a także upředzony 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 uchylbiające 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.*

.....

podpis

# Spis treści

<b>1. Wstęp</b> .....	2
<b>2. Przegląd prac pokrewnych</b> .....	5
<b>3. Baza danych</b> .....	7
<b>4. Wektor parametrów sygnału akustycznego</b> .....	8
<b>5. Metody klasyfikacji</b> .....	10
<b>6. Jakościowa ocena klasyfikacji</b> .....	15
<b>7. Wyniki badań</b> .....	17
<b>8. Podsumowanie</b> .....	25
<b>Bibliografia</b> .....	26

# 1. Wstęp

Mowa i głos człowieka są jednym z podstawowych narzędzi komunikacji międzyludzkiej i wymiany informacji. Poprawna fonacja jest kluczową częścią wielu zawodów, to czy poprawnie i dokładnie mówimy wpływa na odbiór i ocenę naszej wypowiedzi. Błędy w wymowie i choroby narządu głosowego mogą prowadzić do nieścisłości w przekazie informacji i sprawiać dużo kłopotu w codziennym funkcjonowaniu. Choroby krtani są powodem zmian jakości głosu, pierwsze objawy pogorszenia stanu narządu głosowego związane są z szorstkością i chropowatością mowy. Krótkotrwała chrypka może się wiązać z nadużywaniem organu głosu lub zwykłym przeziębieniem. Jednak kiedy szorstkość głosu ma charakter długofalowy i staje się częścią wymowy, to przyczyną staje się dysfunkcja krtani. Co więcej chropowatość głosu czasami jest wczesnym symptomem raka krtani [1]. Dlatego, bardzo ważnym jest, aby zaistniałą dysfunkcję w głosie wykryć jak najszybciej i postawić bezbłędną, obiektywną diagnozę.

Jednym z sposobów diagnozy stanu głosu pacjenta jest jakościowa analiza akustyczna jego wymowy. Wymowa pacjenta jest bardzo ważnym źródłem informacji o stanie funkcjonalnym i anatomicznym krtani. Tradycyjną metodą wykrycia dysfunkcji głosu jest diagnoza lekarska. Na podstawie wymawianych głosek lub fraz przez pacjenta specjalista jest stanie wykryć zaistniałą deformację w wymowie. Postawiona diagnoza jest uzależniona od specjalisty i jego subiektywnej oceny, dlatego należy szukać metody obiektywnej, powtarzalnej i bezbłędnej, która będzie pomocą dla lekarza lub pomoże w automatyzacji detekcji zaistniałych dysfunkcji.

Jednym z badań rozwijanych ostatnimi czasy jest akustyczna analiza nagrania sygnału głosu, która dostarcza fizyczny opis fal dźwiękowych wyemitowanych przez organ wymowy. Wraz z rozwojem cyfrowej rejestracji i metod przetwarzania sygnału mowy możliwe jest tworzenie skutecznej i obiektywnej analizy akustycznej sygnałów audio, która może pomóc w diagnostyce. Wszystkie patologie i choroby ludzkiego układu głosowego wpływają na jakość sygnału mowy pacjenta, dlatego można jednoznacznie rozpoznać aktualny stan określonego źródła głosu. Parametry głosu otrzymane na drodze akustycznej analizy opisują sygnał obiektywnie i powtarzalnie w przeciwieństwie do subiektywnej analizy opartej na percepcji ludzkich narządów odbioru dźwięku. Odpowiednie przetworzenie sygnału fali audio pozwala otrzymać mierzalne charakterystyki ludzkiego głosu, które są podstawą klasyfikacji po-

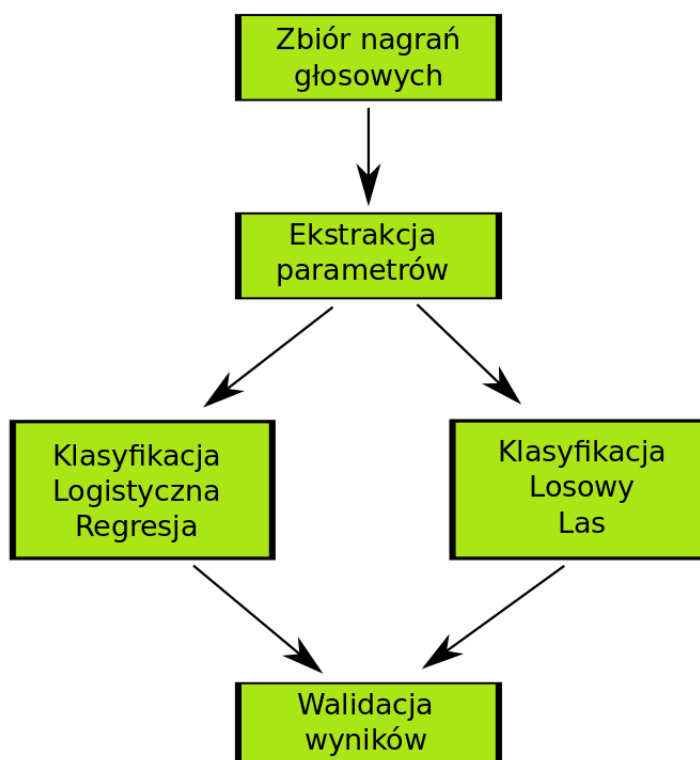
między chorym lub zdrowym przypadkiem, a nawet identyfikacji specyficznej patologii głosu [2]. Podsumowując rejestrując wymowę przedłużonych głosek lub całego zdania można dokonać jakościowej ewaluacji głosu pacjenta. Taka analiza jest początkiem systemu detekcji patologii głosu i prowadzi do finalnej diagnozy.

## Cel pracy

Celem pracy jest przeprowadzenie badań nad systemem detekcji patologii głosu i ocena użyteczności różnych metody akustycznej analizy sygnału mowy w języku Python. Badania sprawdzą informacyjną przydatność wyekstrahowanych wartości charakterystycznych nagrań głosowych w klasyfikacji chory i zdrowych przypadków. Algorytmy systemów uczących się zaimplementowane w języku Python z pomocą biblioteki scikit-learn pozwolą na detekcję patologii głosu w nagraniach audio pacjentów. Wyniki pracy będą zawierać analizę wyników, porównanie skuteczności zastosowanych algorytmów oraz wydajności użytych rozwiązań. Przedstawione rozwiązania w przyszłości mogą być pomocą przy dalszych badaniach lub znajdą zastosowanie w zautomatyzowanych systemach detekcji i diagnozy pacjentów z symptomami patologii głosu. Celem projektu jest kompletny system diagnostyczny pozwalający na obiektywną i dokładną ocenę stanu głosu pacjenta.

## Zakres pracy

Projekt obejmuje kompletny projekt i implementacja systemu, który posłuży do diagnostyki stanu pacjentów. Pierwszy krokiem przed przystąpieniem do detekcji, będzie przedstawienie dyskretnego sygnału audio w postaci wektora parametrów, tak aby umożliwić i zoptymalizować pracę algorytmów systemów uczących się. Wektor parametrów będzie się składał z wartości charakterystycznych dźwięku otrzymanych na drodze przetwarzania sygnału głosu ludzkiego w dziedzinie czasu, częstotliwości i cepstralnej. Analiza będzie się opierała na zbiorze danych zarejestrowanych głosów zdrowych lub chorych pacjentów pobranych z bazy *Saarbruecken Voice Database*. Badaniu poddana zostanie przedłużona głoska /a/ w normalnej tonacji. Wektor parametrów zostanie podzielony na zbiór treningowy i testowy. Zbiór treningowy posłuży optymalizacji dwóch algorytmów uczących się: logistycznej regresji i losowych lasów. W wyniku powstanie aparat pozwalający na dokładną i stosunkowo bezbłędną ocenę stanu krtani pacjenta. Następnie uzyskany system detekcji poddany zostanie jakościowej ocenie opartej na zbiorze testowym, który nie miał udziału w nauczaniu algorytmu. Całkowity proces postępowania podczas tworzenia systemu detekcji patologii mowy przedstawia Rysunek 1.



Rys. 1. Diagram blokowy toku postępowania podczas badań.

## Układ pracy

Praca składa się z dwóch części. Pierwsza część wprowadza czytelnika w ideę i tematykę badań, opisując niezbędną teorię do przeprowadzenia projektu. W rozdziale 2 przedstawiono badania literatury, które stały się inspiracją dla autora. Ich rezultaty są niezwykle istotne, ponieważ pozwalają na dalszą eksploatację terenu automatycznej detekcji patologii głosu i implementację zaproponowanych rozwiązań w języku Python. W następnym rozdziale 3 przedstawiono bazę danych, podług której przeprowadzono dalsze badania. Techniki analizy zbioru nagrań wraz z teorią i wzorami przedstawia rozdział 4. Opis teoretyczny użytych algorytmów oraz ich walidacji został opisany w rozdziale 5. Druga część pracy to praktyczna implementacja opisanych wcześniej rozwiązań. Rozdział 7 przedstawia propozycje i rozwiązania kilku modeli klasyfikacji pacjentów na zdrowych i chorych. Zaprezentowano wizualizację trzech głównych składowych wektora parametrów, tabele z metrykami pozwalającymi ocenić skuteczność algorytmów i wykresy słupkowe przedstawiające wagi użyteczności każdego z parametrów w procesie klasyfikacji.

## 2. Przegląd prac pokrewnych

Automatyzacja obserwacji pacjentów i detekcji laryngologicznych patologii z wykorzystaniem akustycznej analizy głosu zdobywa rosnącą popularność. Wierność nagrań i możliwości przetwarzania cyfrowego sygnału audio zachęca badaczy do rozwoju miarodajnych i jakościowych charakterystyk ludzkiego głosu. Takie metody oferują bardzo dobrą dokładność i obiektywną ocenę stanu chorego. Celem badań jest rzetelne i nieinwazyjne narzędzie, które będzie pomocą dla lekarzy w podjęciu decyzji o stanie w jakim znajduje się krtan pacjenta. Przed przystąpieniem do klasyfikacji zbioru danych należy zadać pytanie czy istnieje różnica pomiędzy akustyczną analizą męskiego, a damskiego głosu. Po odpowiedzi na to zagadnienie warto sięgnąć do pracy [3] lub [4]. Okazuje się, że niektóre parametry charakterystyczne głosu różnią się co do wartości w zależności od płci badanej osoby. Taka informacja jest niezwykle istotna, gdyż zostanie wykorzystana przy tworzeniu wektora parametrów dla algorytmu klasyfikacji lub powstaną dwa różne algorytmy z różnymi wektorami parametrów w zależności od płci.

W celu dokonania poprawnej detekcji stanu głosu pacjenta należy przygotować model zawierający różnego rodzaju wartości charakterystyczne opisujące sygnał dźwięku. Wektor parametrów reprezentuje zbiór nagrań głosowych i jest zbiorem wejściowym dla nauki systemu uczącego się, od jego postaci zależy jakość i dokładność detekcji. Praca [2] opiera się na ocenie istotnych parametrów akustycznych sygnału mowy. W badaniu poddano analizie mapowanie mowy i parametrów w 29-wymiarowej przestrzeni. Parametry mowy zostały wyodrębnione w domenach czasowych, częstotliwościowych i cepstralnych. Algorytm detekcji w pracy [5] jest oparty o zbiór danych zdrowych i chorych pacjentów, gdzie autor ograniczył się do dwóch patologii. Jako reprezentację sygnału autor użył wektora parametrów składającego się z: współczynników jitter i shimmer, podstawowej częstotliwości i 13 współczynników melowych. System detekcji został oparty o naiwny klasyfikator bayesowski. Odmienne podejście proponuje praca [6], gdzie w oparciu o estymację nieliniowych cech dynamiki, przedstawiono automatyczne wykrywanie patologii w układzie fonacyjnym z uwzględnieniem ciągłych zapisów mowy - zależnych od tekstu. Proponowana metodologia jest niezależna od płci i jest również odporna na sygnały o wysokim poziomie patologii, ponieważ nie wymaga oceny częstotliwości podstawowej. Autor pracy [7] wybrał wektor cech charakterystycznych oparty na 31 paramet-

trach wydobytych z dziedziny czasu, częstotliwości i cepstralnej. Następnie dokonał stratnej kompresji poprzez algorytm Principal Component Analysis do 17 najbardziej wartościowych informacyjnie i mających największą czułość na deformacje mowy parametrów. W artykule [4] autor opisuje poszczególne metody akustycznej analizy i przedstawia porównanie narzędzi automatycznie analizujących wybrany sygnał audio.

Ostatnim etapem budowy poprawnie działającego systemu detekcji jest wybór odpowiedniego algorytmu klasyfikacji. W ostatnich latach techniki systemów uczących się i eksploracji danych przyniosły wiele dokładny i bardzo dobrze działających rozwiązań [8] lub [9]. Te potężne narzędzia można tak wykorzystać, aby poprowadziły do lepszych osiągnięć w działaniu detekcji patologii głosu. Autor [10] korzystając z zbioru danych Saarbruecken Voice Database (SVD) analizuje 1410 przypadków pacjentów. W celu klasyfikacji na zdrowych i chorych, porównuje wykorzystanie dwóch algorytmów klasyfikacji: k-Means i Random Forest, gdzie ten drugi osiąga dokładność wynoszącą niemal 100%. Dodatkowo artykuł przedstawia porównanie klasyfikacji z kompresją wektora parametrów: PCA/kPCA i bez kompresji, gdzie wektor zawiera 28 parametrów.



### 3. Baza danych

Zbiór danych tworzą nagrane i udostępnione przez Instytut Fonetyki w Saarland. Nagrania są dostępne *online* na stronie *Saarbruecken Voice Database* (SVD) [11]. SVD zawiera nagrania głosu ponad 2000 przypadków. Każdy pacjent opisany jest poprzez:

- nagrania wymowy trzech głosek /a/, /i/ oraz /u/ w tonacjach niskiej, normalnej i wysokiej,
- nagrania wymowy trzech głosek /a/, /i/ oraz /u/ w rosnąco-malejącej tonacji,
- nagrania wymowy sentencji „*Guten Morgen, wie geht es Ihnen?*”.

Każde z nagrań przechowywane jest w postaci mowy i sygnału EGG. Obydwa sygnały mogą być wyeksportowane w oryginalnej formie NSP i EGG, jak również w formie WAV. Czas trwania nagrań samogłosek wynosił od 1 do 4 sekund. Wszystkie nagrania zostały zarejestrowane z częstotliwością próbkowania 50 kHz i z rozdzielczością 16 bitów. Baza SVD składa się z nagrań mowy pacjentów cierpiących łącznie na 71 różnych chorób narządu głosu. Z bazy usunięto nagrania uszkodzone lub o czasie trwania krótszym niż 1 sekunda. W badaniach wykorzystano nagrania samogłoski /a/ dla normalnej tonacji w formie WAV. Z powodu wyraźnej różnicy w zachowaniu głosu w zależności od płci dla mężczyzn i kobiet, nagrania głosowe zostały przeanalizowane dla każdej z płci osobno.

## 4. Wektor parametrów sygnału akustycznego

Pierwszym etapem przy konstrukcji algorytmu detekcji, jest wstępna transformacja sygnału mowy, tak aby otrzymać zbiór parametrów, na podstawie którego zostanie dokonana diagnoza pacjenta. Rejestracja sama w sobie i wstępne przetwarzanie sygnału audio nie czyni go w pełni użytecznym dla procesu identyfikacji i oceny zmian deformacji i patologii. Dlatego staje się konieczne opracowanie i opis zarejestrowanych testów fonetycznych przy użyciu zestawu parametrów, które następnie posortowane w odpowiedniej strukturze - wektorze parametrów, pozwolą opracować modele opisu deformacji w mowie. Takie modele mogą być podstawą procesu rozpoznawania i oceny zmian patologicznych w głosie pacjenta. Wektor parametrów dla każdego z nagrań będzie stanowił zbiór wejściowy i będzie służył optymalizacji systemów uczących się w celu klasyfikacji pacjentów.

W celu ilościowej oceny procesu fonacji konieczny jest jego opis parametryczny. Sygnał przeanalizowano w dziedzinie czasu, częstotliwości i cepstralnej. W pracy rozważono następujące parametry tego procesu: wartość maksymalną i minimalną, wartość średnią kwadratową (RMS), kurtozę, współczynnik skośność, częstotliwość podstawową (F0), współczynnik szybkości przejścia przez zero (ang. *Zero-crossing Rate*), i 10 współczynników mel-cepstralnych (MFCC). Do oszacowania głośności wykorzystywana jest wartość średnia kwadratowa (RMS) sygnału [1](#). Kurtoza jest miarą płaskości sygnału, a współczynnik skośności jest miarą asymetrii rozkładu prawdopodobieństwa zmiennej losowej o wartości rzeczywistej. Kurtuza i współczynnik skośności zostały obliczone w języku Python dzięki bibliotece SciPy [\[12\]](#).

$$RMS_x = \sqrt{\sum_{n=1}^N x(n)^2} \quad (1)$$

Częstotliwość podstawową obliczono na podstawie położenia wartości maksymalnej w dziedzinie częstotliwości. Współczynnik szybkości przejścia przez zero jest szybkością zmian znaków wzdłuż sygnału, tj. szybkości, z jaką sygnał zmienia się z dodatniego na ujemny lub z powrotem. Współczynniki MFCC są szeroko stosowane w rozpoznawaniu mowy, ponieważ odzwierciedlają dobrze odbiór dźwięku poprzez zwiększenie słyszalnej częstotliwości i są mniej wrażliwe na hałas [\[13\]](#). MFCC zostały zaprojektowane, aby odzwierciedlać naturalną reakcję systemu słuchowego na stymulację mową.

Po ekstrakcji cech charakterystycznych sygnału głosu, wektor parametrów został poddany standaryzacji według wzoru 2. Standaryzacja parametrów jest dla wielu systemów uczących się ważnym krokiem wstępnego przetworzenia zbioru danych. Standaryzacja polega na przeskalowaniu wektora cech, tak aby ich rząd wielkości był taki sam. W wypadku braku normalizacji parametrów algorytm może zostać optymalizowany tylko po kątem parametrów, które mają większe wartości, czego wynikiem będzie błędna predykcja.

$$x = \frac{x - \mu}{\sigma} \quad (2)$$

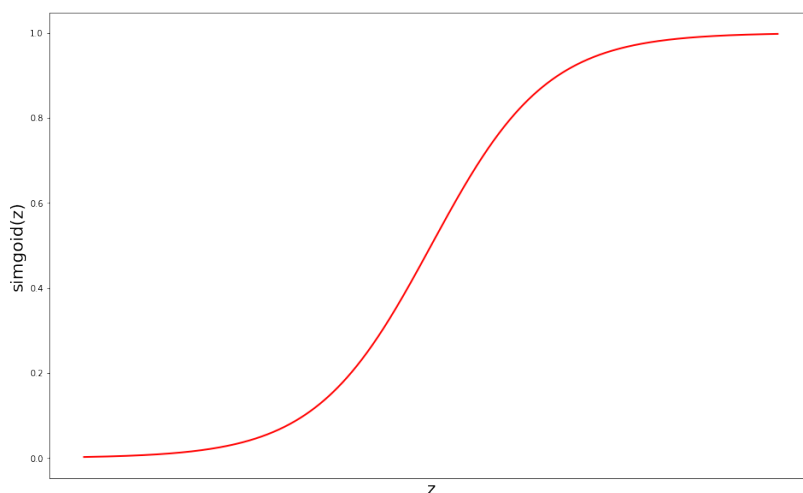
## 5. Metody klasyfikacji

### Regresja logistyczna

Regresja logistyczna może być mylącą nazwą, ponieważ model służy liniowej klasyfikacji, a nie regresji. Regresja logistyczna wykorzystywana jest do klasyfikacji. W naszym modelu klasyfikacja ogranicza się do podziału na dwie klasy - chory lub zdrowy pacjent. Aby uzyskać binarną klasyfikację 0 zdrowy przypadek, 1 chory przypadek, wzór na liniową regresję **3** (gdzie  $X$  to wektor parametrów nagrania głosu każdego z pacjentów,  $\theta$  to zbiór wag dla wektora  $X$ ) należy poddać funkcji sigmoidalnej nazwanej funkcją logistyczną **4**. Dzięki czemu model nie będzie przewidywał prawdopodobieństw, które są mniejsze od 0 lub większe od 1, Rysunek **2**. Model logistycznej regresji jest zaprojektowany tak, aby przypisywać prawdopodobieństwo pomiędzy 0% a 100% na to, że etykieta - wartość wyjściowa należy do rzeczywistej klasy, w naszym przypadku stanu pacjenta.

$$Z = \theta^T X \quad (3)$$

$$\sigma(Z) = \frac{1}{1 + \exp^{-Z}} \quad (4)$$



**Rys. 2.** Funkcja sigmoidalna ograniczająca wartości wyjściowe przykładowej liniowej regresji do przedziału  $[0,1]$ .

W rezultacie otrzymamy hipotezę  $h(X)$  5, która informuje nas o prawdopodobieństwie wystąpienia wyniku wyjściowego 1, w naszym przypadku będzie to chory pacjent 6. Równanie 6 możemy przekształcić do postaci 7, która przyjmuje postać funkcji logitowej.

$$h(X) = \sigma(\theta^T X) \quad (5)$$

$$h_\theta(X) = P(y = 1|X; \theta) = 1 - P(y = 0|X; \theta) = \frac{1}{1 + \exp^{-\theta^T X}} \quad (6)$$

$$\log\left(\frac{P(y = 1|X; \theta)}{1 - P(y = 1|X; \theta)}\right) = \theta^T X \quad (7)$$

Aby znaleźć wektor wag  $\theta$  dla wektora parametrów nagrań głosowych pacjentów  $X$  należy zoptymalizować model regresji logistycznej. Optymalizacja polega na odnalezieniu minimum funkcji kosztu  $J(\theta)$  8, gdzie  $m$  to liczba pozycji, w naszym przypadku będzie to liczba pacjentów, a  $y$  to rzeczywista etykieta - stan pacjenta.

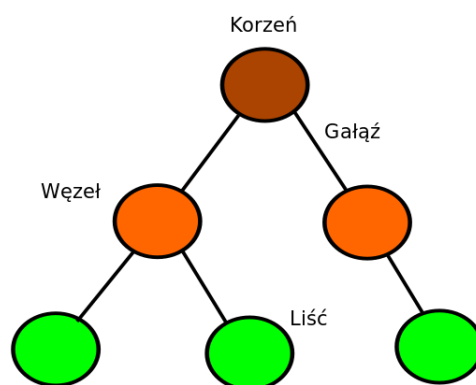
$$J(\theta) = \frac{1}{m}(-y^T \log(h) - (1 - y)^T \log(1 - h)) + \frac{\lambda}{2m} \theta^T \theta \quad (8)$$

Pierwszy człon wzoru 8 informuje o rozbieżności pomiędzy predykcją modelu, a rzeczywistą etykietą. Drugi człon po znaku plus to regularyzacja. Współczynnik regularyzacji  $\lambda$  wpływa na stopień generalizacji modelu, mniejsze wartości precyzują silniejszą regularyzację, co sprawnie zapobiega przeuczeniu się algorytmu dla parametrów różniących się od siebie w małym stopniu. Minimalizację funkcji kosztu przeprowadzamy za pomocą gradientowego zejścia według wzoru 9, gdzie  $\alpha$  odpowiada za wielkość kroku gradientu.

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \quad (9)$$

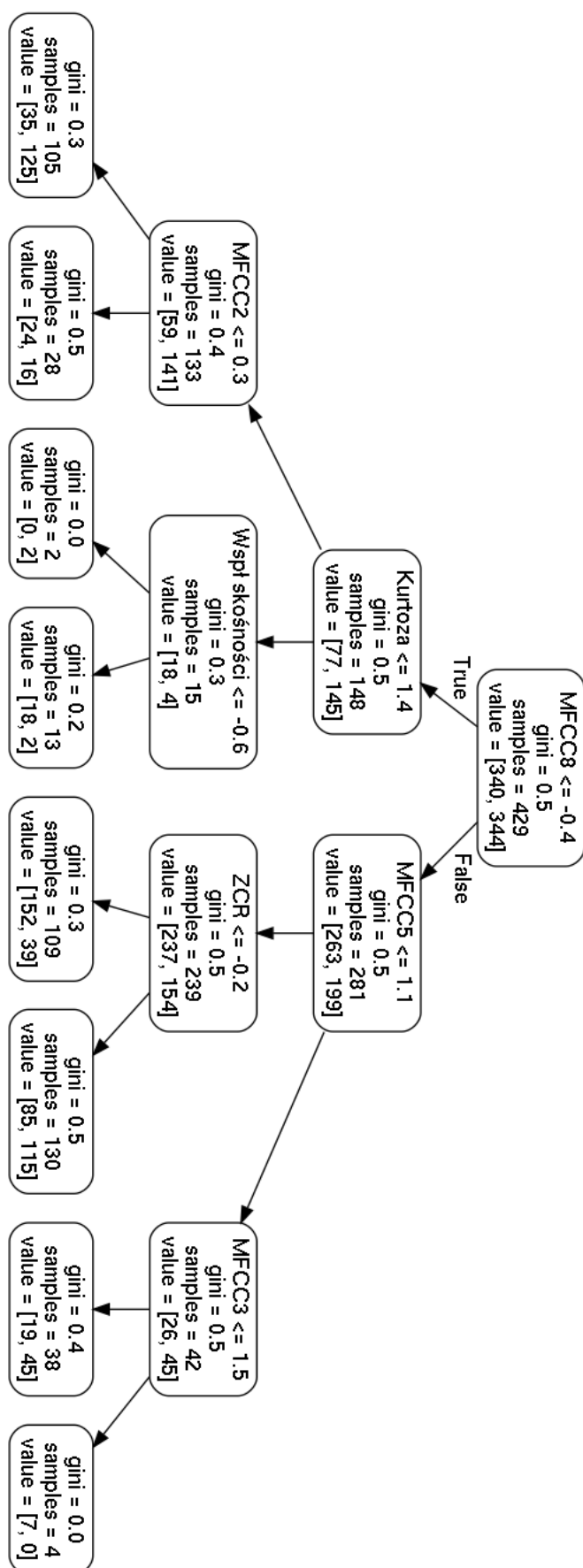
## Lasy losowe (ang. *Random Forest*)

Aby w pełni rozumieć model losowego lasu, najpierw należy wprowadzić pojęcie drzewa decyzyjnego, podstawowej części tworzącej losowy las. Algorytm drzew decyzyjnych opiera się na strukturze drzew binarnych (Rysunek 3) i jest wielopoziomym procesem decyzyjnym.



**Rys. 3.** Przykładowa struktura drzewa binarnego

Każdy węzeł drzewa decyzyjnego zawiera kryterium dotyczącego danej obserwacji, kryterium jest tak sformułowane, aby można było odpowiedzieć Prawda lub Fałsz. Co więcej, dla każdej Prawdziwej i Fałszywej odpowiedzi istnieją oddzielne gałęzie. Jako kryterium podziału może być stosowana miara informacji wyrażona jako indeks Gini lub entropia wzajemna. Bez względu na odpowiedzi na pytania, w końcu osiągamy prognozę zawartą w liściach drzewa decyzyjnego. Podsumowując zaczynając od korzenia drzewa oraz odpowiadając na napotkane pytania po drodze otrzymujemy predykcję. Podczas szkolenia przekazujemy modelowi wszelkie dane, które są istotne dla dziedziny problemowej (parametry opisujące nagrania głosowe) oraz prawdziwą wartość, którą chcemy, aby model mógł się nauczyć przewidzieć. Model uczy się wszelkich zależności między danymi, a wartościami, które chcemy przewidzieć. Drzewo decyzyjne tworzy strukturę wybierając najlepsze kryteria, które należy sformułować, tak aby umożliwić możliwie najdokładniejsze oszacowanie stanu pacjenta, Rysunek 4. Kiedy poprosimy drzewo decyzyjne o klasyfikację stanu pacjenta, musimy podać dane tak samo spreparowane jak te, które wykorzystano podczas szkolenia. Wyjściem drzewa decyzyjnego jest prognoza oparta na strukturze, której się nauczyło.



**Rys. 4.** Przykładowa struktura zredukowanego drzewa decyzyjnego, będącego jednym z wielu elementów losowego lasu. Diagram otrzymany przy pomocy pakietu scikit-learn [14].

Podstawową ideą lasu losowego jest połączenie wielu drzew decyzyjnych w jeden model. Każde drzewo decyzyjne uwzględnia losowy podzbiór parametrów podczas formułowania kryteriów i ma dostęp do losowego, ograniczonego zbioru danych treningowych. Prognozy podejmowane przez indywidualne drzewa decyzyjne mogą nie być dokładne i w dużym stopniu różnić się od siebie wariancją. W momencie predykcji losowy las oblicza średnią wszystkich oszacowań poszczególnych drzew decyzyjnych. W celu ewaluacji algorytmu klasyfikacji zbiór wejściowy jest dzielony na dwa podzbiory: zbiór treningowy oraz zbiór OOB (ang. *out-of-bag*). Na podstawie zbioru OOB można określić ile elementów zbioru treningowego zostało błędnie sklasyfikowanych. Lasy losowe nie wymagają wiedzy eksperckiej i są odporne na nadmierne dopasowanie do danych - przetrenowanie.

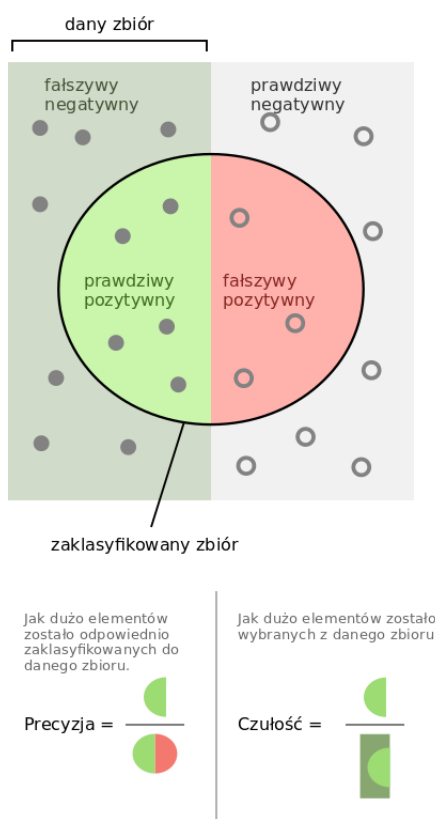


## 6. Jakościowa ocena klasyfikacji

Podstawową metryką ewaluacji modelu klasyfikacji jest dokładność. Dokładność informuje o tym jaką część predykcji nasz model poprawnie sklasyfikował **10**.

$$\text{dokładność} = \frac{\text{Liczba poprawnych predykcji}}{\text{Całkowita liczba predykcji}} \quad (10)$$

Kiedy pracujemy na zbiorze danych o niezbalansowanym podziale klas, to sama dokładność nie jest wystarczającą metryką dla pełnej oceny modelu. Znacząca dysproporcja w liczbie pozytywnych, a negatywnych etykiet jest bardzo częstym przypadkiem w zbiorach danych na medycznym obszarze. Rozwiązaniem jest zastosowanie dwóch dodatkowych metryk precyzji i czułości. Rysunek **5** przedstawia ich graficzną zależność. Grafika została zaczerpnięta z [15] i zedytowana według własnych potrzeb.



**Rys. 5.** Precyzja i czułość

## Czułość

Czułość lub zdolność modelu do znalezienia wszystkich istotnych przypadków w zbiorze danych, jest to liczba prawdziwych pozytywów podzielona przez liczbę prawdziwych pozytywów plus liczbę fałszywych negatywów **11**. Prawdziwe pozytywy to punkt danych zaklasyfikowany jako pozytywny przez model, który faktycznie jest pozytywny (co oznacza, że są poprawne), a fałszywe negatywy są punktami danych, które model identyfikuje jako negatywne, które faktycznie są pozytywne (niepoprawne). W przypadku patologii głosu prawdziwymi pozytywami są prawidłowo określone chore, a fałszywe negatywy są osobami które zostały zaklasyfikowane jako zdrowe, a faktycznie są chore. Czułość, to zdolność modelu do znalezienia wszystkich interesujących punktów w zbiorze danych.

$$\text{czułość} = \frac{\text{prawdziwe pozytywy}}{\text{prawdziwe pozytywy} + \text{fałszywe negatywy}} \quad (11)$$

## Precyzja

Precyzja **12** jest definiowana jako liczba prawdziwych pozytywów podzielona przez liczbę prawdziwych pozytywów plus liczba fałszywych pozytywów. Fałszywe pozytywy to przypadki, w których model błędnie określa jako pozytywne, które są w rzeczywistości negatywne lub w naszym przykładzie osoby, które model uznaje za chore, a które nie są. Podczas, gdy czułość wyraża możliwość znalezienia wszystkich istotnych punktów w zestawie danych, precyzja wyraża proporcję punktów danych, które nasz model zaklasyfikowała poprawnie, a ich całkowitą ilość prognoz.

$$\text{precyzja} = \frac{\text{prawdziwe pozytywy}}{\text{prawdziwe pozytywy} + \text{fałszywe pozytywy}} \quad (12)$$

## Wynik „F<sub>1</sub>”

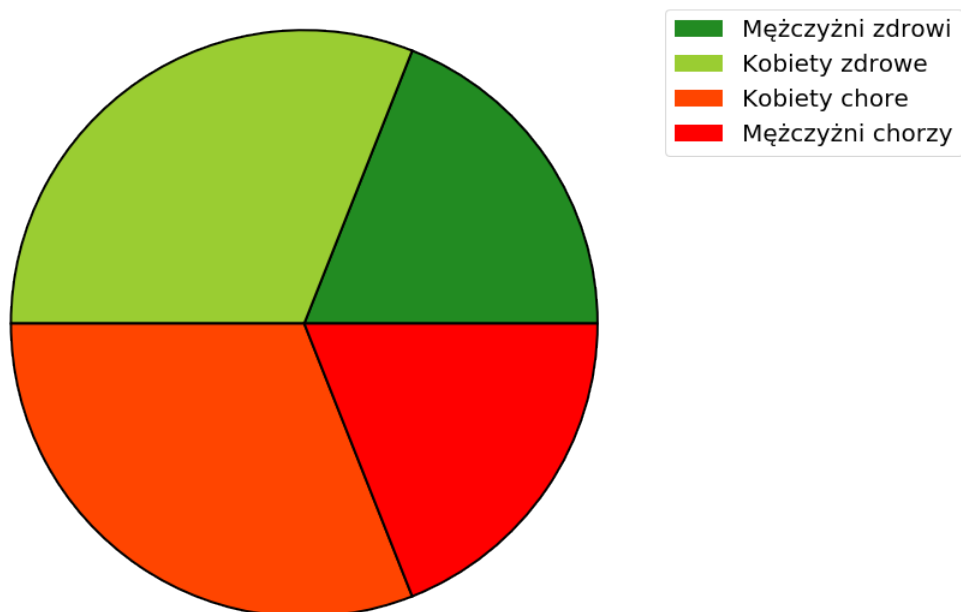
W przypadkach, w których chcemy znaleźć optymalną równowagę w precyzji i czułości, możemy połączyć te wskaźniki za pomocą tak zwanego wyniku F<sub>1</sub>. Wynik F<sub>1</sub> jest średnią harmoniczną precyzji i czułości **13**.

$$F_1 = 2 * \frac{\text{czułość} * \text{precyzja}}{\text{czułość} + \text{precyzja}} \quad (13)$$

Jeśli chcemy znaleźć zbalansowany model klasyfikacji z optymalną równowagą pomiędzy czułością, a precyzją należy dążyć do maksymalizacji wyniku F<sub>1</sub>.

## 7. Wyniki badań

Ten etap badań ma na celu zaprezentowanie projektów systemów i ich weryfikacji. Systemy umożliwią klasyfikację pacjentów na zdrowych i chorych. U badanych występują różne schorzenia mowy. System przeprowadza klasyfikację binarną pacjentów: 1 wykryto patologię, a 0 pacjent zdrowy. Jako bazę danych wykorzystałem zbiór nagrań głosowych pacjentów o nazwie *Saarbruecken Voice Database* (SVD). Szczegółowy opis bazy SVD znajduje się w Rozdziale 3. Swoje badania ograniczyłem do nagrań głosek /a/ w normalnej tonacji, wybrałem tylko jedną głoskę w jednej tonacji. Swój wybór argumentuję przeglądem podobnych badania w Rozdziale 2, gdzie nie zauważyłem diametralnej różnicy w wektorze parametrów i rezultatach detekcji deformacji w mowie. Dlatego uznałem, że dla wstępnej i szybkiej weryfikacji zaprezentowanych rozwiązań wystarczy jedna głoska w normalnej tonacji, jeśli dany system osiągnie zadowalające rezultaty, to jego ocena jakościowa zostanie oparta o wszystkie głoski w trzech możliwych tonacjach. Aby skonstruować zbiór danych wybrałem 428 zdrowych i 428 chorych nagrań głosów kobiet, u mężczyzn liczby wyniosły 259 zdrowych i 259 chorych przypadków, Rysunek 6.



**Rys. 6.** Postać zbioru danych z uwzględnieniem płci i stanu pacjenta

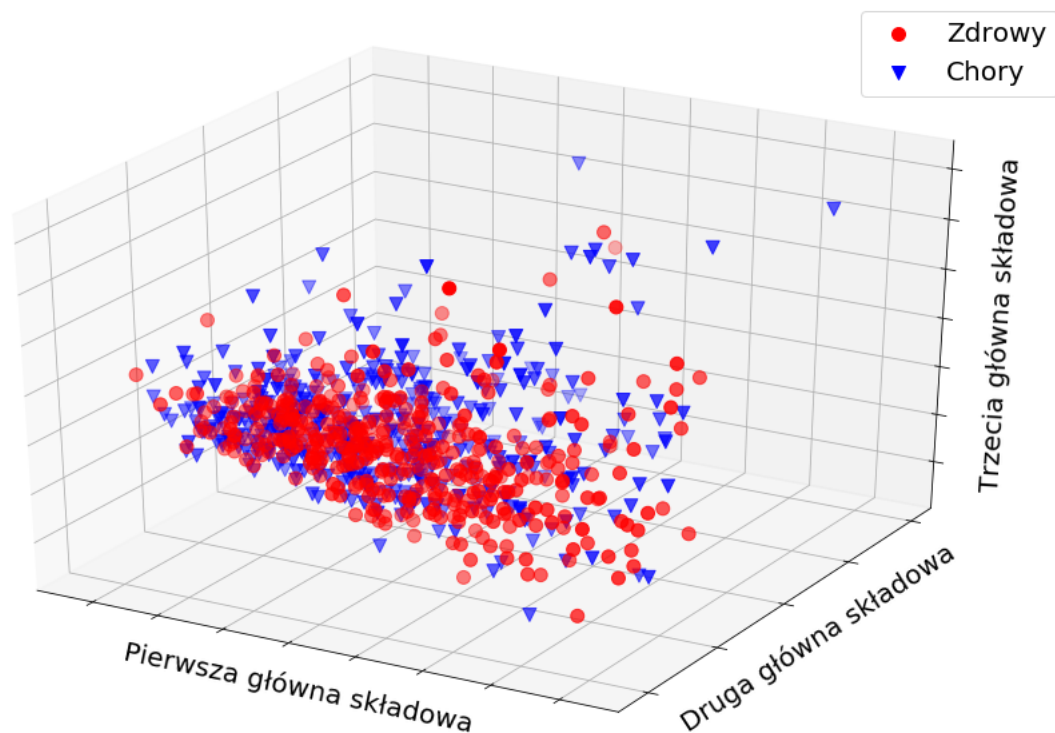
Stosunek zdrowych pacjentów do chorych wyniósł 1 do 1, to znaczy 50% stanowiła grupa osób o poprawnej fonacji, a pozostała grupa zawierała osoby z różnymi deformacjami głosu. Każda z płci została potraktowana indywidualnie i powstały dwa oddzielne systemy w zależności o płci pacjenta. Aby mieć pewność do wyników klasyfikacji i mieć porównanie wybrałem z biblioteki scikit-learn [14] dwa różne działające algorytmy: logistycznej regresji 5 i metody lasów losowych 5. Zaprojektowałem i opracowałem propozycje zbiorów parametryczny nagrań fonetycznych głosu, które umożliwiły detekcję zniekształceń sygnału głosu. Ponieważ parametry różniły się zakresem zmienności oraz wartością, wektor parametrów poddałem standaryzacji opisanej wzorem 2. Zbiór danych opisany poprzez wektor parametrów został podzielony w stosunku 80 do 20, to znaczy 80% danych wykorzystano do treningu systemów uczących się, a pozostałe 20% danych posłużyło za test wytrenowanych algorytmów.

Z punktu widzenia klasyfikacji ważne jest, aby parametry pozwoliły na oddzielenie różnych klas od siebie. Dlatego, aby mieć wizualny pogląd na wektor parametrów, dokonałem przekształcenia wektora parametrów do trzech głównych składowych. Innymi słowy skompresowałem wielowymiarowe wektory do trzech wymiarów, a następnie stworzyłem na ich podstawie wykresy. Kolejnym krokiem był trening algorytmów systemów uczących, których wyniki zostały przedstawione w tabelach. Ponadto jednym atrybutów wyjściowych metody losowych lasów jest wektor wag przypisanych dla każdego z parametrów w czasie treningu. Taki wykres jest niezwykle istotny, ponieważ pozwala wybrać parametry o najwyższej wartości informacyjnej, co oszczędzi czas obliczeń. Dodatkowo można uniknąć przetrenowania w przypadku parametrów o niskim wpływie na klasyfikację, poprzez ich wychwycenie i usunięcie. Wynik prognozy porównany z zestawem treningowym posłużył za jakościową analizę klasyfikacji, szczegółowy opis znajduje się w Rozdziale 6.

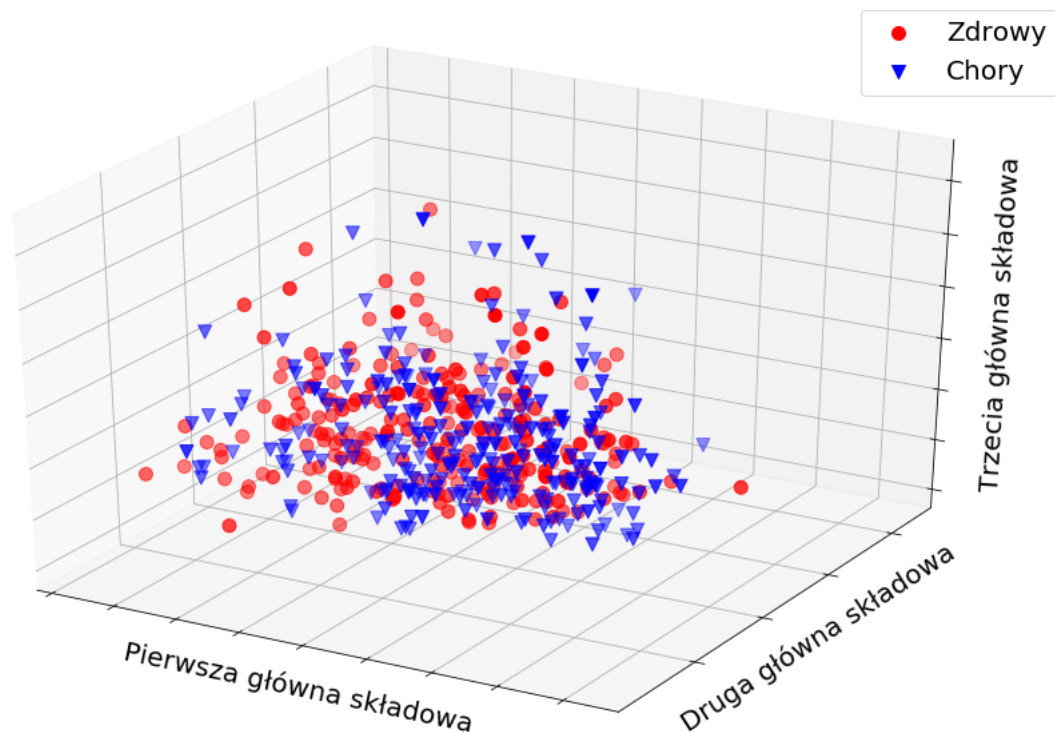
Swoje badanie oparłem na metodzie kolejnych prób, co pozwoliło mi na wprowadzanie usprawnień do już działających rozwiązań, a następnie ich weryfikację z poprzednim wynikiem detekcji.

## I próba

Moim celem na wstępie było zaprojektowanie prostego i szybkiego systemu detekcji, tak aby przekonać się o jego skuteczności, a na jego podstawie opracować nowy i bardziej szczegółowy algorytm. Dlatego ograniczyłem się do analizy sygnału audio w postaci w dziedzinie czasu, w rezultacie otrzymałem wektor złożony z 5 parametrów: wartości minimalnej i maksymalnej, kurtozy, współczynnika skośności (skos) i wartość średnio kwadratowej sygnału (RMS). Rysunki 7 i 8 przedstawiają analizę głównych składowych parametrów z rozróżnieniem na stan krtani pacjenta z podziałem na kobietę i mężczyznę.



**Rys. 7.** Prezentacja 3 głównych głównych składowych samogłoski /a/ w normalnej tonacji dla kobiety z rozróżnieniem pomiędzy chorym, a zdrowym przypadkiem.



**Rys. 8.** Prezentacja 3 głównych składowych wektora parametrów samogłoski /a/ w normalnej tonacji dla mężczyzny z rozróżnieniem pomiędzy chorym, a zdrowym przypadkiem.

Na rysunkach nie można zaobserwować wyraźnej granicy w rozkładzie pomiędzy zdrowych i chorych pacjentów. Nasuwa się wniosek, że wektory parametrów nie zawierająca wystarczającej wagi informacyjnej w postaci wartości charakterystycznych dla nagrań głosów zdrowych lub chorych. W wyniku czego system detekcji oparty na algorytmach systemów uczących się nie będzie w stań poprawnie przeprowadzić klasyfikacji stanu pacjenta.

Pomimo braku wyraźnej granicy w rozkładzie i reguły rządzącej wartości parametrów dla pacjentów zdrowych i chorych warto sprawdzić czy algorytmy detekcji rzeczywiście osiągają słabe rezultaty. Pozwoli to na porównanie ze sobą jakości parametrów i odrzucenia tych o niskiej przydatności dla systemu. Ponadto tak czynność pozwoli na weryfikację, czy algorytmy ulegają przetrenowaniu (ang. *overfitting*). Tabele 1 i 2 prezentują precyzję, czułość, wynik F1 i dokładność dwóch algorytmów: logistycznej regresji i losowego lasu dla kobiet i mężczyzn głoski /a/ w normalnej tonacji.

**Tab. 1.** Precyzja, czułość, wynik F1 i dokładność algorytmu logistycznej regresji dla kobiet i mężczyzn głoski /a/ w normalnej tonacji.

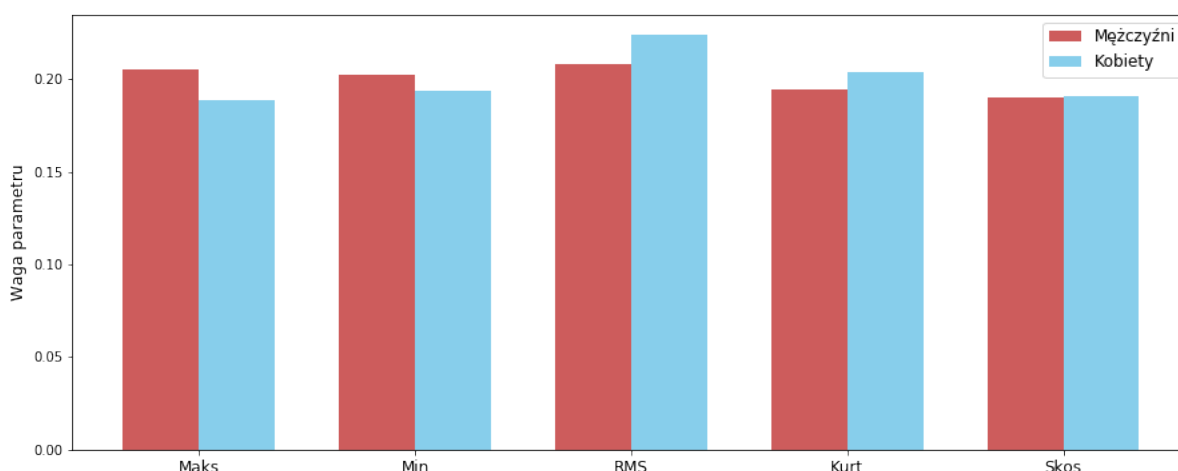
płeć	stan pacjenta	precyzja	czułość	wynik F1	dokładność
kobieta	zdrowy	0.57	0.65	0.61	0.57
	chory	0.56	0.48	0.52	
mężczyzna	zdrowy	0.54	0.51	0.52	0.51
	chory	0.48	0.51	0.50	

**Tab. 2.** Precyzja, czułość, wynik F1 i dokładność algorytmu losowych lasów dla kobiet i mężczyzn głoski /a/ w normalnej tonacji.

płeć	stan pacjenta	precyzja	czułość	wynik F1	dokładność
kobieta	zdrowy	0.61	0.64	0.63	0.60
	chory	0.59	0.57	0.58	
mężczyzna	zdrowy	0.58	0.56	0.57	0.56
	chory	0.53	0.55	0.54	

Najwyższy wyniki F1 osiąga algorytm losowych drzew dla kobiet i wnosi on odpowiednio 63% dla poprawnego wykrycia zdrowego i 58% dla chorego stanu pacjenta. Natomiast najgorszy wynik F1 przedstawia algorytm logistycznej regresji przy klasyfikacji mężczyzn na zdrowych 52% i chorych 50%. Wyniki potwierdzają hipotezę wysnutą na podstawie Rysunków 7 i 8. Systemy detekcji patologii nie są wystarczająco wytrenowane (ang. *underfitting*) należy zmodyfikować wektor parametrów wejściowych, aby poprawić ich skuteczność. Wartość przydatność parametrów w czasie nauki algorytmu losowych drzew przedstawia Rysunek ???. Nie zważyłem wyraźnej różnicy informacyjnej pomiędzy cechami charakterystycznymi zbioru nagrań, dlatego wszystkie parametry zostaną użyte w następnej próbie.

**Rys. 9.** Wagi dla każdego z parametrów przypisane przez algorytm losowych lasów w zależności od płci pacjenta.

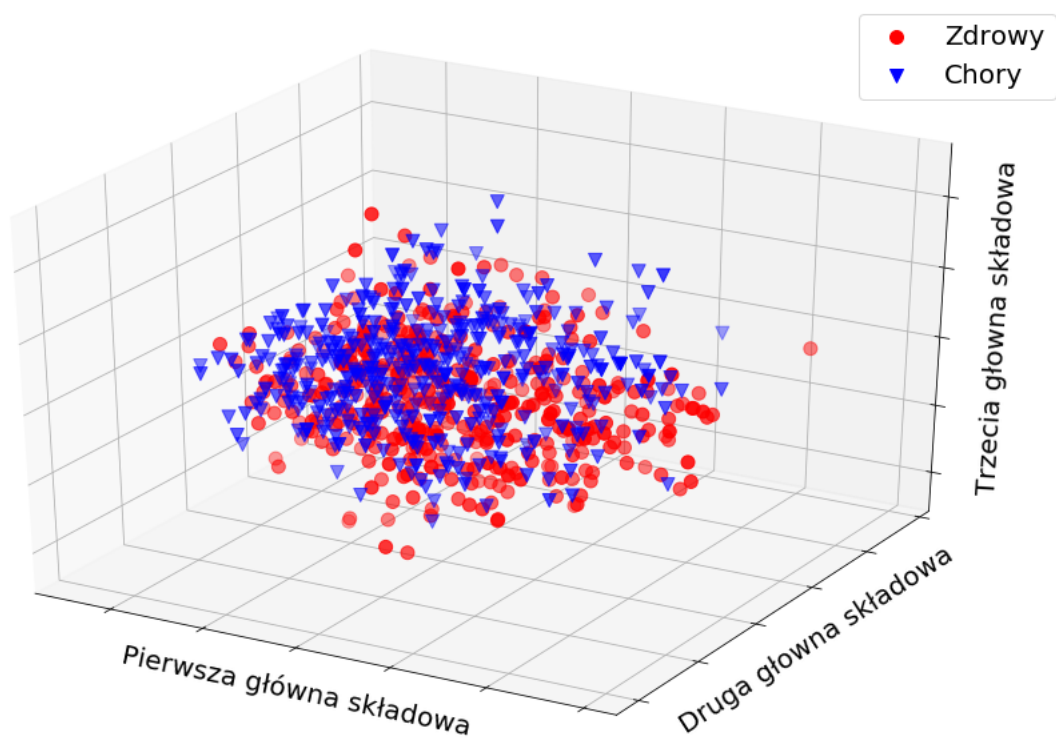


Brak wyraźnej granicy pomiędzy zdrowym, a chorym przypadkiem przekłada się na małą efektywność algorytmów. Potwierdzenie tej hipotezy pozwoliło wykazać, że algorytmy nie ulegają przetrenowaniu. Wektor parametrów zawiera dane, których rozkład nie wskazuje jednoznacznie na osobę zdrową lub chorą. Parametry otrzymane z analizy sygnału w dziedzinie czasu nie są wystarczającym źródłem do przeprowadzenia klasyfikacji. Powyższy proces pozwolił na szybki test zaprojektowanego modelu i opracowanie dalszych kroków usprawnienia systemu detekcji patologii mowy. W kolejnej próbie skupię się na opracowaniu wektora parametrów o większej przydatności w procesie detekcji. Wektor parametrów zostanie uzupełniony o analizę sygnału w dziedzinie częstotliwościowej i cepstralnej.

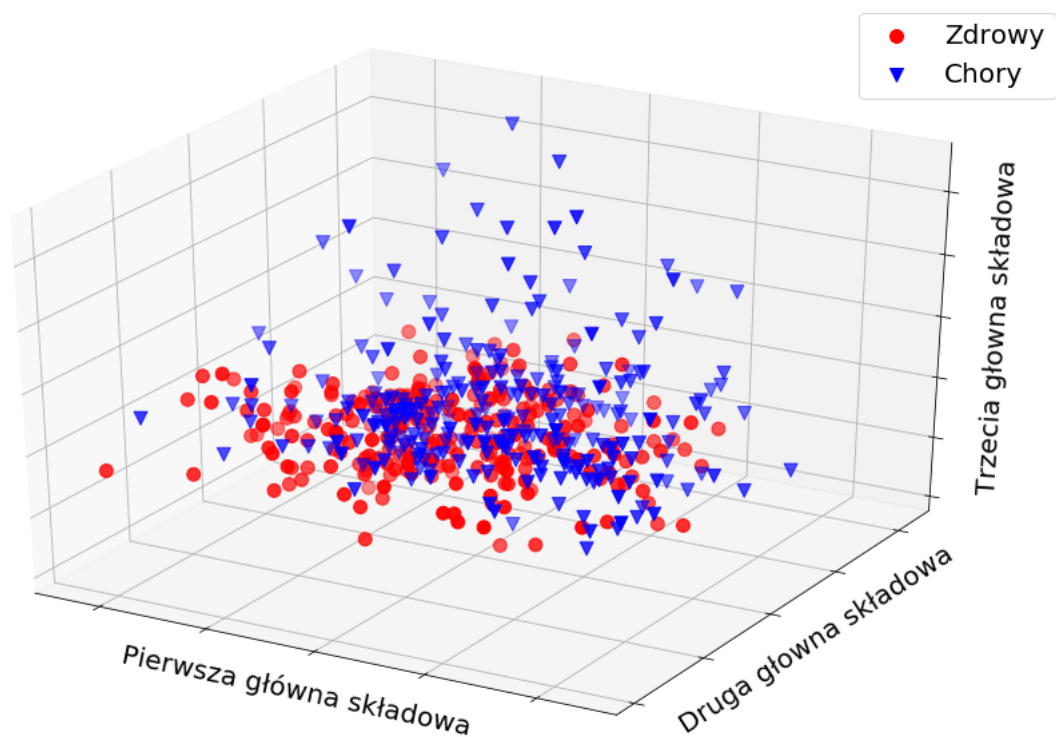
## II próba

Poprzednia próba dowiodła, że analiza sygnału nagrań pacjentów w dziedzinie czasu nie jest wystarczająca dla poprawnej klasyfikacji, dlatego wektor uzupełnię o analizę sygnału w dziedzinie częstotliwościowej i cepstralnej. Nowy wektor parametrów będzie się składał z 17 parametrów: wartości minimalnej i maksymalnej, kurtozy, współczynnika skośności (skos) i wartość średnio kwadratowej sygnału (RMS), częstotliwości podstawowej (F0), współczynnika szybkości przejścia przez zero (ang. *Zero-crossing Rate*, ZRC), i 10 współczynników mel-cepstralnych (MFCC). W języku Python obliczenie współczynników MFCC i współczynnika ZRC umożliwia biblioteka LibROSA [16]. Rysunki 10 i 11 przedstawiają analizę trzech głównych składowych wektora parametrów z rozróżnieniem na stan krtani pacjenta z podziałem na kobietę i mężczyznę. Pojawia się lekka tendencja w zależności od stanu zdrowia pacjenta dla mężczyzn jak i kobiet.





**Rys. 10.** Prezentacja 3 głównych głównych składowych samogłoski /a/ w normalnej tonacji dla kobiety z rozróżnieniem pomiędzy chorym, a zdrowym przypadkiem.



**Rys. 11.** Prezentacja 3 głównych składowych wektora parametrów samogłoski /a/ w normalnej tonacji dla mężczyzny z rozróżnieniem pomiędzy chorym, a zdrowym przypadkiem.



Aby sprawdzić jakość nowego wektora 17 parametrów reprezentujących zbiór nagrań sygnałów audio w dziedzinie czasu, częstotliwości i cepstrum przeprowadziłem klasyfikację pacjentów przy pomocy logistycznej regresji i losowego lasu. Tabele ?? i ?? prezentują precyzję, czułość, wynik F1 i dokładność dwóch algorytmów.

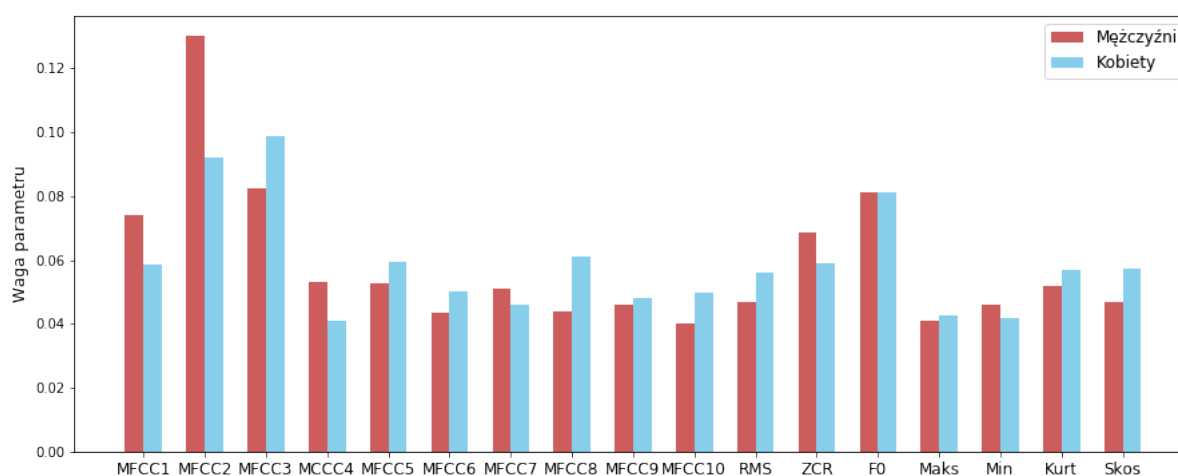
**Tab. 3.** Precyzja, czułość, wynik F1 i dokładność algorytmu logistycznej regresji dla kobiety i mężczyzny głoski /a/ w normalnej tonacji.

płeć	stan pacjenta	precyzja	czułość	wynik F1	dokładność
kobieta	zdrowy	0.73	0.69	0.71	0.72
	chory	0.70	0.74	0.72	
mężczyzna	zdrowy	0.76	0.55	0.69	0.70
	chory	0.66	0.83	0.71	

**Tab. 4.** Precyzja, czułość, wynik F1 i dokładność algorytmu losowych lasów dla kobiet i mężczyzny głoski /a/ w normalnej tonacji.

płeć	stan pacjenta	precyzja	czułość	wynik F1	dokładność
kobieta	zdrowy	0.71	0.70	0.71	0.70
	chory	0.70	0.71	0.71	
mężczyzna	zdrowy	0.76	0.63	0.73	0.73
	chory	0.69	0.81	0.75	

Obydwa algorytmu uzyskały bardzo podobne wyniki. W porównaniu z I próbą można zaobserwować wyraźną poprawę w klasyfikacji. Najgorszy wyniki F1 osiągnął algorytm logistycznej regresji w klasyfikacji mężczyzn na zdrowych 69% i chorych 71%, a najlepszy algorytm losowych drzew w klasyfikacji mężczyzn zdrowych 73% i chorych 75%. Rysunek 12 przedstawia wektor wag przypisanych dla każdego z parametrów w czasie treningu algorytmu losowych lasów. Wagi parametrów różnią się i zależą płci pacjenta.



**Rys. 12.** Wagi dla każdego z parametrów przypisane przez algorytm losowych lasów w zależności od płci pacjenta.

Najważniejsze parametry to MFCC2, MFCC3 i F0. Parametry otrzymane na drodze analizy sygnału w dziedzinie częstotliwościowej i cepstralnej odznaczają się większą przydatnością, niż te otrzymane z ekstrakcji w dziedzinie czasu. Żaden z parametrów nie wykazuje się wystarczająco małą wagą, aby uznać go za zbędny.

Powiększenie wektora parametrów nowymi wielkościami charakterystycznymi wyraźnie poprawiło jakość z jaką algorytm dokonuje klasyfikacji. Wraz z dodawaniem nowych parametrów pojawia się tendencja do coraz lepszej generalizacji systemu w polu detekcji patologii mowy. Uzupełnienie wektora parametrów o cechy opisane w pracach przedstawionych w Rozdziale 2 pozwoli na uzyskanie algorytmu o wyższej dokładności w klasyfikacji pacjentów na zdrowych i chorych. Warto skupić się na cechach opisujących sygnał w dziedzinie częstotliwościowej i cepstralnej, gdyż posiadają najwyższe wagi podczas treningu algorytmu losowych drzew. Podczas dwóch prób algorytm losowych lasów osiągał lepsze lub bardzo podobne wyniki do algorytmu logistycznej regresji. Dlatego losowe lasy należy wybrać jako finalny system klasyfikacji pacjentów na zdrowych i chorych, tak jak miało to miejsce w [10].

## 8. Podsumowanie

Jakość życia może zależeć od stanu naszego systemu głosowego, zwłaszcza gdy jest to jeden z głównych sposobów komunikacji w naszym codziennym życiu. W literaturze zaproponowano wiele rodzajów projektowania systemów diagnostyki narządów głosowych, których wybrane pozycje zostały przedstawione w Rozdziale 2. Jednocześnie bardzo mała liczba prac i badań implementuje swój system w języku Python. Praca zawiera szereg badań i poszukiwań efektywnych rozwiązań w temacie detekcji patologii głosu w tymże języku programowania.

Największym wyzwaniem podczas pracy w technice przetwarzania mowy jest opracowanie nowych sposobów opisu sygnału w postaci parametrów, które będą charakterystyczne dla danego sygnału. Pozwoli to wydobyć z nagrań głosowych pacjenta użyteczne informacje o stanie jego mowy. Celem badań było ustalenie w jakim stopniu zmiany w sygnale mowy mogą zostać wykorzystane jako źródło informacji diagnostycznej, prognostycznej i kontrolnej przy rozwiązywaniu problemów medycznych związanych z chorobami laryngologicznymi. Zaproponowane rozwiązania mogą wesprzeć lekarza w sformułowaniu diagnozy lub posłużyć automatyzacji procesu detekcji patologii mowy.

Praca przedstawia różne podejścia do projektowania i implementacji kompletnych systemów klasyfikacji w języku Python. System podejmuje próbę jednoznacznego podziału pacjentów na zdrowych i chorych w oparciu o wielowymiarowe wektory cech charakterystycznych opisujący ludzki głos. W tym celu wykorzystano 1 374 nagrań samogłosek /a/ o normalnej tonacji. Zbiór danych zaczerpnięto z bazy Saarbruecken Voice Database (SVD), szerzej opisany w Rozdziale 3. Zaprojektowano model klasyfikacji binarnej, w wyniku której system wskazywał, czy osoba jest zdrowa czy nie. Zaproponowana metodologia pozwala ocenić stan zdrowia pacjenta, bez względu na postęp choroby narządu głosu. System potrafi diagnozować wiele rodzajów zaburzeń głosu w różnych stanach postępu choroby i może być podstawowym narzędziem do wykrywania możliwych chorób głosowych u pacjentów poddawanych badaniom przesiewowym głosu.

Przedstawiono i przetestowano kilka wersji wektorów parametrów wyekstrahowanych z sygnału audio w dziedzinie czasu, częstotliwości i cepstralnej. Ostateczna wersja zaproponowanego systemu klasyfikacji jest złożona z wektora 17 parametrów przedstawia wynik F1 w okolicach 70% dla obu płci. Badanie przedstawiają możliwe

rozwiązania wraz ich oceną jakościową. Podczas badań zauważono niewielki odsetek implementacji systemów detekcji patologii mowy w języku Python. Dodatkowo język Python nie posiada dedykowanych narzędzi lub bibliotek do dokładnej analizy akustycznej sygnału audio pod kątem detekcji patologii mowy. Pracę można rozwinąć o nowe parametry opisane w pracach przedstawionych w Rozdziale 2. Uzupełnienie wektora parametrów o nowe wartości charakterystyczne otrzymane na drodze analizy akustycznej nagrań zgodnie z badaniami przeprowadzonymi w Rozdziale 7 mogą poprawić jakość klasyfikacji pacjentów na zdrowych i chorych. Szczególną uwagę należy zwrócić na parametry opisujące sygnał w dziedzinie częstotliwościowej i cepstralnej, gdyż posiadają najwyższe wagi podczas treningu systemów uczących się, Rysunek 12. W przypadku ewaluacji modelu o satysfakcjonujących wynikach, finalny model detekcji patologii mowy należy rozszerzyć o analizę wszystkich trzech przedłużonych głosek: /a/, /i/, /u/ w trzech tonacjach wraz z metrykami opisującymi ich skuteczność. Wszystkie algorytmy zaimplementowane podczas prac nad badaniami dostępne są w moim repozytorium na stronie GitHub [17].

# Bibliografia

- [1] J. C. Stemple, N. Roy, B. K. Klaben, *Clinical voice pathology: Theory and management*. Plural Publishing, 2014.
- [2] Z. W. Engel, M. Kłaczyński, W. Wszolek, "A vibroacoustic model of selected human larynx diseases," *International Journal of Occupational Safety and Ergonomics*, vol. 13, no. 4, s. 367–379, 2007.
- [3] J. P. Teixeira P. O. Fernandes, "Jitter, shimmer and hnr classification within gender, tones and vowels in healthy voices," *Procedia Technology*, vol. 16, s. 1228–1237, 2014.
- [4] M. Brockmann, M. J. Drinnan, C. Storck, P. N. Carding, "Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task," *Journal of voice*, vol. 25, no. 1, s. 44–53, 2011.
- [5] M. Dahmani M. Guerti, "Vocal folds pathologies classification using naïve bayes networks," in *Systems and Control (ICSC), 2017 6th International Conference on*. IEEE, 2017, s. 426–432.
- [6] J. R. Orozco, J. F. Vargas, J. B. Alonso, M. A. Ferrer, C. M. Travieso, P. Henriquez, "Voice pathology detection in continuous speech using nonlinear dynamics," in *Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on*. IEEE, 2012, s. 1030–1033.
- [7] D. Panek, A. Skalski, J. Gajda, "Quantification of linear and non-linear acoustic analysis applied to voice pathology detection," in *Information Technologies in Biomedicine, Volume 4*. Springer, 2014, s. 355–364.
- [8] C. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag New York, 2006, vol. 1.
- [9] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1.
- [10] D. Hemmerling, A. Skalski, J. Gajda, "Voice data mining for laryngeal pathology assessment," *Computers in biology and medicine*, vol. 69, s. 270–276, 2016.

- [11] B. Woldert-Jokisz, "Saarbruecken voice database," 2007.
- [12] Scipy. The SciPy community. [Online]. Available: <https://docs.scipy.org/doc/scipy/reference/index.html>
- [13] R. Loughran, J. Walker, M. O'Neill, M. O'Farrell, "The use of mel-frequency cepstral coefficients in musical instrument identification," in *International Computer Music Conference*. International Computer Music Association, 2008.
- [14] scikit-learn. Machine Learning in Python. [Online]. Available: <https://scikit-learn.org/stable/index.html>
- [15] Walber. (2014) Precision and recall. [Online]. Available: <https://commons.wikimedia.org/wiki/File:Precisionrecall.svg>
- [16] Librosa. Librosa Development Team. [Online]. Available: <http://librosa.github.io/librosa/index.html#librosa>
- [17] B. Tynski. Voice pathology detection in python. [Online]. Available: <https://github.com/tynski/voicePathology>