Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej

Katedra Metrologii i Elektroniki

Praca dyplomowa inżynierska

Wykorzystanie systemów uczących się do detekcji patologii głosu

The use of machine learning in voice pathology detection

Autor: Bartosz Tyński
Kierunek studiów: Elektrotechnika
Opiekun pracy: dr inż. Mirosław Socha

Kraków, 16 stycznia 2019

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 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 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.

. . . . . .

podpis

Spis treści

1. Wstęp

Mowa i głos człowieka są jednym z podstawowych narzędzi komunikacji miedzyludzkiej 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 [1]. 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. Dlatego, ważne 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 deformacje w wymowie. Jednakże postawiona diagnoza jest uzależniona od specjalisty i jego subiektywnej oceny, a dokładność detekcji nie jest stała i zależy od wielu zmiennych. 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 jednoznaczne 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-

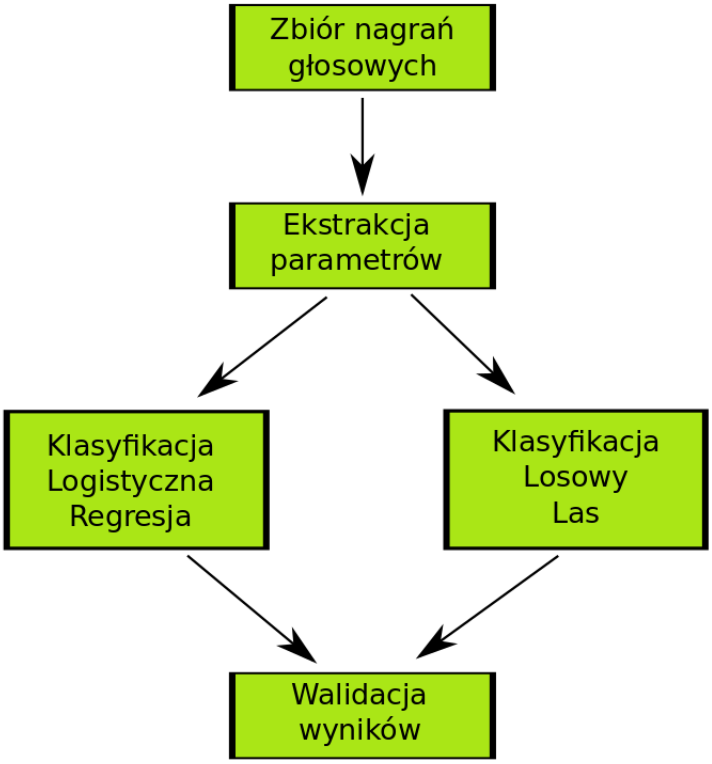
między chorym lub zdrowym przypadkiem, a nawet identyfikacji specyficznej patologii głosu [2]. Podsumowując rejestrując wymowę przedłużonych głosek lub całego zdania można dokonać jakościowej ewaluacji głosu pacjenta. Taka analiza jest początkiem systemu detekcji patologii głosu i prowadzi do finalnej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

Cel pracy

Celem pracy jest przeprowadzenie badań nad systemem detekcji patologii głosu i ocena użyteczności różnych metod akustycznej analizy sygnału mowy. Badania sprawdzą informacyjną przydatność wyekstrahowanych wartości charakterystycznych nagrań głosowych w klasyfikacji nagrań głosowych pacjentów na zdrowych i chorych. Algorytmy systemów uczących się zaimplementowane w języku Python z pomocą biblioteki *scikit-learn [3]* pozwolą na detekcję patologii głosu w nagraniach audio pacjentów. Wyniki pracy będą zawierać analizę, 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

Niniejsza praca obejmuje projekt kompletnego systemu wraz z implementacją. Opracowany system posłuży do diagnostyki stanu pacjentów. Pierwszy krokiem przed przystąpieniem do klasyfikacji pacjentów, 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 poddane zostaną przedłużone głoski /a/, /i/, /u/ 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 ocenę stanu krtań 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 patologi mowy przedstawia Rys. 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 pozwalają na dalszą eksploatację terenu automatycznej detekcji patologii głosu i implementację zaproponowanych rozwiązań w języku *Python*. Rozdział 3. przedstawia 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. Część praktyczna zakończona jest oceną jakościową zaproponowanego systemu w odniesieniu od trzech przedłużonych głosek /a/, /i/, /u/ w normalnej tonacji.

2. Przegląd prac pokrewnych

Automatyzacja obserwacji pacjentów i detekcji laryngologicznych patologii z wykorzystaniem akustycznej analizy głosu zdobywa rosną 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 krtań pacjenta. Przed przystąpieniem do klasyfikacji zbioru danych należy zadać następujące pytanie. Czy istnieje różnica pomiędzy akustyczną analizą męskiego, a damskiego głosu?

Po odpowiedź na to zagadnienie warto sięgnąć do pracy [4] lub [5]. Okazuje się, że niektóre parametry charakterystyczne głosu, takie jak *jitter* lub *shimmer* różnią się co do wartość 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 algorytmu z różnymi wektorami parametrów w zależności od płci.

W celu dokonania poprawnej detekcji stanu głosu pacjenta należy przygotować model zawierający rożnego 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 [6] jest oparty o zbiór danych zdrowych i chory 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 [7], 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.

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]). Zaawansowane algorytmy systemów uczących się takie jak: *Random Forest* lub *Convolutional Neural Network* 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%.

3. Wykorzystana 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 [12]. Obydwa sygnały mogą być wyeksportowane w oryginalny formacie EGG, jak również w formacie WAV [13]. Czas trwania nagrań samogłosek wynosił od 1. do 4. sekundy. Wszystkie nagrania zostały zarejestrowane z częstotliwości 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 trzech przedłużonych głosek /a/, /i/, /u/ w normalnej tonacji w formacie WAV. Przeanalizowano 1 374 nagrań pacjentów. Pochodzą one od 856 kobiet, z których 428 było zdrowych, a 428 chorych oraz 518 mężczyzn, z których 258 było zdrowych, a 258 należało do grupy osób chorych. 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. Wektor parametrów to zbiór wartości charakterystycznych każdego z nagrań głosowych, jakość wektora parametrów świadczy, o tym jak dobrze potrafi zobrazować konkretne nagranie. Zależy nam aby otrzymać obraz jak najpełniejszy, ponieważ na podstawie wektora parametrów 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 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. W pracy rozważono parametry otrzymany w procesie analizy sygnał w trzech dziedzinach:

- czasu - wartość maksymalną i minimalną, wartość średnią kwadratową (RMS), kurtozę, współczynnik skośność,
- częstotliwości - częstotliwość podstawową (F_0),
- cepstralnej - 10 współczynników mel-cepstralnych (MFCC), współczynnik szybkości przejścia przez zero (ang. *Zero-crossing Rate*).

Dziedzina czasu

Do oszacowania głośności wykorzystywana jest wartość średnia kwadratowa (RMS) sygnału (1):

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

Kurtoza jest miarą płaskości sygnału (2):

$$Kurt_x = \frac{\frac{1}{n}\sum(x - \mu)^4}{\sigma^4} \tag{2}$$

Czym większa jest kurtoza, tym mniejsze jest spłaszczenie rozkładu, czyli większa koncentracja wokół wartości oczekiwanej. Współczynnik skośności jest miarą asymetrii rozkładu prawdopodobieństwa zmiennej losowej o wartości rzeczywistej (3):

$$\gamma_1 = \frac{\frac{1}{n}\sum(x - \mu)^3}{\sigma^3} \tag{3}$$

Dziedzina częstotliwości

Periodyczność w mowie obejmuje zdolność do generowania ciągłego przepływu powietrza podczas produkcji samogłosek o przedłużonej fonacji. Stabilność i niezmienność takiego przepływu przez struny głosowe może zostać scharakteryzowana parametrycznie poprzez zmienność amplitudy lub częstotliwości. Odcinek czasu pomiędzy kolejnymi zwarciami fałdów głosowych wyznacza najmniejszą powtarzającą się sekwencję w sygnale mowy. Odcinek ten nazywany jest okresem podstawowym. Odwrotność tego okresu definiuje częstotliwość podstawową (F_0). F_0 jest jednym z ważniejszych parametrów charakteryzujących źródło mowy. Jego wartość zależy m.in. od płci, wieku oraz stanu zdrowotnego mówcy. Dla kobiet wartość F_0 mieści się w przedziale 160-960 Hz, a dla mężczyzn 80-480 Hz [14]. Na podstawie parametru F_0 otrzymanego w efekcie badań głosek typu: /a/, /i/, /u/ o przedłużonej fonacji możliwe jest wykrycie anomalii struktury anatomicznej, które występują w różnych stanach patologicznych [15]. Do wyznaczenia parametru F_0 można wykorzystać kilka metod, w przedstawionych badaniach częstotliwość podstawową obliczono na podstawie położenia wartości maksymalnej w dziedzinie częstotliwości. Do zmiany dziedziny czasu na dziedzinę częstotliwościową wykorzystano algorytm szybkiej transformaty Fouriera (FFT) [16]. Po wykonaniu FFT sygnału odszukano wartość maksymalną transformaty. Częstotliwość odpowiadająca położeniu maksimum jest poszukiwaną F_0.

Dziedzina cepstralna

Obecnie klasyczna analiza spektralna metod sygnałów głosowych jest często uzupełniana o metody takie jak liniowa analiza predykcyjna, analiza falkowa lub homomorficzna w dziedzinie cepstrum [2], [17]. Cepstrum określa się jako odwrotną transformatę Fouriera (IFFT) [16] logarytmu dziesiętnego widma częstotliwościowego sygnału. Równanie (4) opisuje wyznaczenie cepstrum rzeczywistego:

$$Cepstrum = IFFT(log_{10}(|FFT(x)|)) \tag{4}$$

Cepstrum daje lepszy obraz struktury sygnału harmonicznego i pozwala na separację istniejącego hałasu w przekształconym sygnale [18]. Wielu autorów podkreśla wagę czynników cepstralnych w diagnostycznej ocenie zmian patologicznych w mowie [19]. Dlatego analiza sygnału mowy w sposób podobny do zjawiska słyszenia, wymaga przekształcenia skali częstotliwości do skali cepstralno-melowej. Odpowiada ono subiektywnemu wrażeniu wysokości dźwięku. Po transformacji do skali melowej, obliczane jest cepstrum. Proces wyznaczenia współczynników mel-cepstralnych (MFCC) składa się z dwóch etapów: pre-emfazy i okienkowania . Celem pre-emfazy jest wzmocnienie składowych o wysokiej częstotliwości i osłabienie składowych o niskiej częstotliwości. Wzmocnienie wysokich częstotliwości sygnału mowy sprawia, że sygnał jest odporny na zakłócenia pochodzące z otoczenia. Natomiast procedura okienkowania dzieli sygnał na odcinki czasu z określoną długością [14]. Dzięki czemu sygnał akustyczny zachowuje stacjonarny charakter [20]. Współczynniki MFCC są szeroko stosowane w rozpoznawaniu mowy, ponieważ odzwierciedlają dobrze odbiór dźwięku poprzez zwiększenie słyszalnej częstotliwości i są mniej wrażliwe na hałas [21]. MFCC zostały zaprojektowane, aby odzwierciedlać naturalną reakcję systemu słuchowego na stymulację mową.

Standaryzacja cech

Po ekstrakcji cech charakterystycznych sygnału głosu, wektor parametrów został poddany standaryzacji. Standaryzacja parametrów jest dla wielu systemów uczących się ważnym krokiem wstępnego przetworzenia zbioru danych. Polega na przeskalowaniu wektora cech, tak aby ich rząd wielkości był taki sam, równanie (5):

$$x = \frac{x - \mu}{\sigma} \tag{5}$$

W naszym wypadku od każdego z parametrów odejmowana zostaje wartość średnia tego parametru wszystkich nagrań, a następnie wynik zostaje podzielony przez odchylenie standardowe parametru dla wszystkich nagrań. Jeśli nie wykonamy standaryzacji algorytm może zostać optymalizowany tylko po kątem parametrów, które mają większe wartości, czego wynikiem będzie błędna predykcja.

5. Metody klasyfikacji

W kolejnym kroku należy przygotować algorytm, który będzie klasyfikował dane wejściowe. Klasyfikacja ma za zadanie określić stan pacjenta na podstawie przeanalizowanego i opracowanego wcześniej wektora parametrów. Aby móc dokonać klasyfikacja wybrany algorytm klasyfikacji należy wytrenować. W tym celu wykorzystywany jest zbiór danych uczących, gdzie w naszych badaniach będzie to wektor cech charakterystycznych, przedstawionych i opisanych w Rozdziale 4. Wytrenowany algorytm będzie zwracał przynależność do danej klasy na podstawie danych wejściowych. Nasze badania zajmują się klasyfikacją binarną, gdzie wyróżniane są dwie klasy. Taka klasyfikacja pozwoli na podział pacjentów na zdrowych lub chorych.

Regresja logistyczna

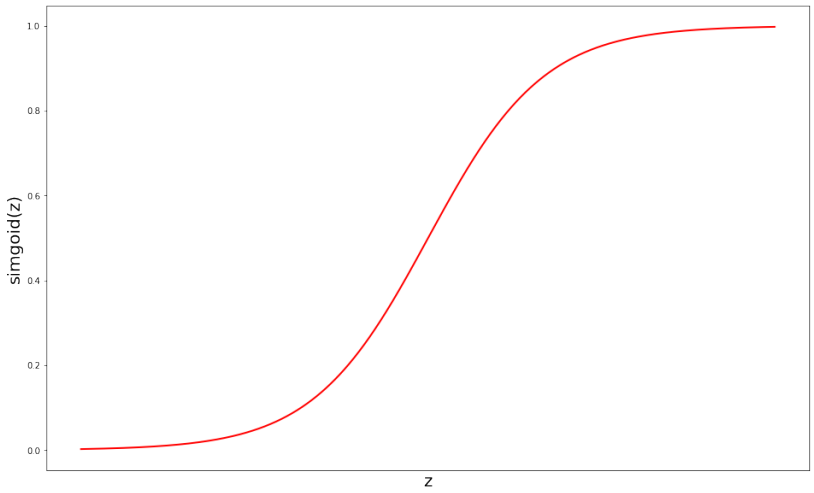
Regresja logistyczna może być mylącą nazwą, ponieważ model służy liniowej klasyfikacji, a nie regresji. W naszym modelu klasyfikacja ogranicza się do binarnego podziału na dwie klasy: zdrowy przypadek i chory przypadek. Jednakże, aby to zrobić najpierw należy wprowadzić wzór regresji liniowej (6):

$$Z = \theta^T X \tag{6}$$

Wektor X zawiera parametry nagrania głosu każdego z pacjentów, natomiast θ to zbiór wag dla wektora X. Mając wektor Z, czyli zbiór wartość otrzymany jako rezultat wzoru (6) zależy nam, aby otrzymane wartości reprezentowały dwie klasy (zdrowy/chory). Jednym z narzędzi matematycznych umożliwiających taką transformację jest funkcja sigmoidalna, wzór (7):

$$\sigma(Z) = \frac{1}{1 + \exp^{-Z}} \tag{7}$$

Dzięki jej wykorzystaniu nasz model nie będzie przewidywał wartości, które są mniejsze od 0 lub większe od 1, Rys. 2. Dodatkowo ze względu na rozkład funkcji (7) znacznie prościej jest przeprowadzić klasyfikację binarną.



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

Łącząc ze sobą wprowadzone wcześnie wzory (6) i (7), otrzymujemy hipotezę $h(X)$ (8), czyli model logistycznej regresji.

$$h\left(X\right) = \sigma\left(\theta^T X\right) \tag{8}$$

Dla logistycznej regresji hipoteza (8) zwraca prawdopodobieństwo wystąpienia wyniku wyjściowego 1., w naszym przypadku będzie to chory pacjent (9).

$$h_\theta\left(X\right) = P\left(y=1|X;\theta\right) = 1 - P\left(y=0|X;\theta\right) = \frac{1}{1+\exp^{-\theta^T X}} \tag{9}$$

Podsumowując podając na wejście funkcji *h(X)* wektor parametrów *X* dla danego pacjenta, otrzymujemy dla prognozę jego stanu. Wynik prognozy w głównej mierze zależy od wektora wag θ przez który przemnażany jest wektor wejściowy *X*. Aby dostać jak najdokładniejszą prognozę funkcja *h(X)* musi posiadać taki wektor wag θ,aby wartość zwrócona przez funkcję kosztu *J(θ)* (10) była jak najmniejsza.

$$J(\theta) = \frac{1}{m}(-y^T log(h) - (1-y)^T log(1-h)) + \frac{\lambda}{2m}\theta^T\theta \tag{10}$$

Wartość m to liczba pozycji, w naszym przypadku będzie to liczba pacjentów, a y to rzeczywista etykieta, czyli stan pacjenta. Pierwszy człon wzoru (10) informuje o rozbieżności pomiędzy predykcją modelu, a rzeczywistą etykietą. Drugi człon po znaku plus to regularyzacja. Współczynnik regularyzacji λ 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.

Celem treningu logistycznej regresji jest znaleźć taką hipotezę *h(X)*, dla której funkcja kosztu *J(θ)* przyjmie jak najmniejszą wartość. Aby znaleźć minimum funkcji kosztu *J(θ)* należy znaleźć optymalny wektor wag θ dla logistycznej regresji. Jed-

nym z podejść do optymalizacji algorytmu logicznej regresji jest minimalizacja funkcji kosztu za pomocą gradientowego zejścia (11).

$$\theta = \theta - \alpha \nabla J(\theta) \tag{11}$$

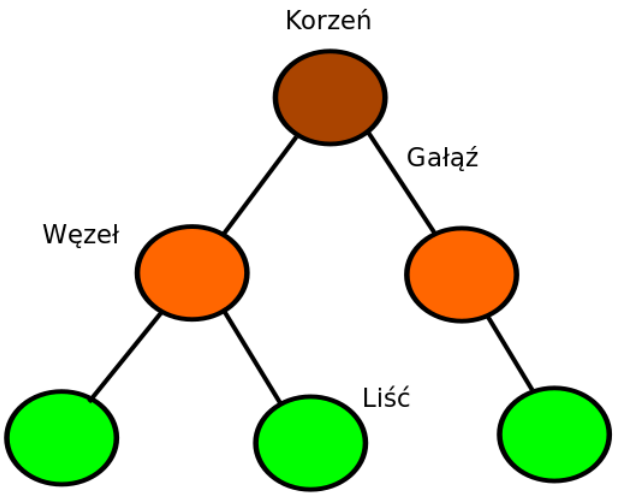
Współczynnik α (ang. *learing rate*) odpowiada za wielkość kroku gradientu. Im większy współczynnik α tym szybciej można dotrzeć do minimum, ale zbyt duża szybkość uczenia wiążę się z ryzykiem ominięcia wartości minimalnej i w rezultacie niestabilność algorytmu. Gradient dla funkcji kosztu $\nabla J(\theta)$ można opisać wzorem (12):

$$\nabla J(\theta) = \frac{1}{m} X^T (h(X) - y) \tag{12}$$

Podsumowując, aby wytrenować algorytm logistycznej regresji należy przeprowadzić optymalizację hipotezy $h(X)$ (9) pod kątem gradientowego zejścia (12). Polega ona na aktualizacji wartości wektora wag θ według wzoru (11) i obliczaniu wartość funkcji kosztu $J(\theta)$ (10). Jeśli osiągniemy satysfakcjonujące rezultaty w postaci wystarczająco małego błędu predykcji kończymy trening algorytmu.

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 (Rys. 3.) i jest wielopoziomowym procesem decyzyjnym.

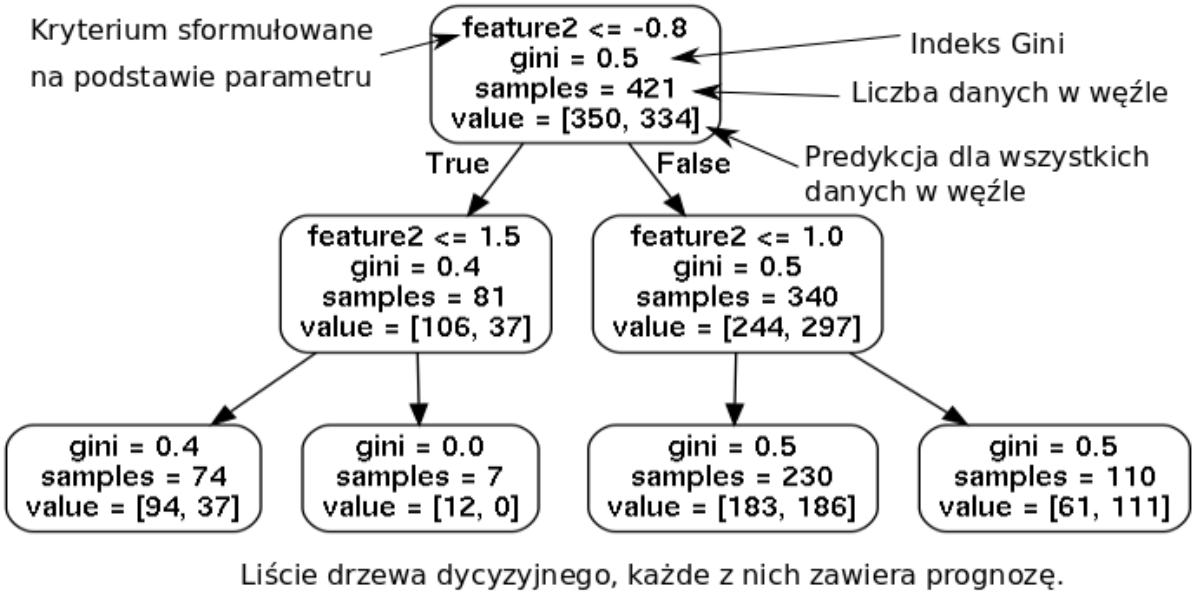


Rys. 3. Przykładowa struktura drzewa binarnego

Każdy węzeł drzewa decyzyjnego zawiera kryterium, które jest sformułowane na podstawie wektora parametrów. Na każde takie kryterium można jasno odpowie-

dzieć Prawda lub Fałsz. Dla każdej Prawdziwej i Fałszywej odpowiedzi istnieją oddzielne gałęzie, które prowadzą do kolejnego węzła. Jeśli zaczniemy od korzenia drzewa oraz odpowiemy na kryteria napotkane w węzłach, będziemy podążać w dół drzewa decyzyjnego, aż natrafimy na węzeł bez gałęzi, czyli liść. Każdy z liści zawiera prognozę opartą na odpowiedziach do kryteriów, które znajdują się w drodze do niego.

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, czyli takie które mają największy indeks ważności Gini [22]. Aby drzewo decyzyjne dokonało klasyfikacji 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. Przykład struktury i zasady drzewa decyzyjnego przedstawia Rys. 4.



Rys. 4. Przykładowa struktura zredukowanego drzewa decyzyjnego, będącego jednym z wielu elementów losowego lasu.

Znając strukturę i zasadę działania drzewa decyzyjnego możemy wprowadzić pojęcie lasu losowego. 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. Takie podejście pozwoli na konstrukcję drzew decyzyjnych opartych na różnych podzbiorach danych wejściowych, w rezultacie otrzymamy szeroki wachlarz prognoz opartych na różnych strukturach drzew. Prognozy podejmowane przez indywidualne drzewa decyzyjne mogą nie być dokładne i w du-

żym stopniu różnić się od siebie. W momencie predykcji losowy las oblicza średnią wszystkich oszacowań poszczególnych drzew decyzyjnych i na tej podstawie dokonuje klasyfikacji.

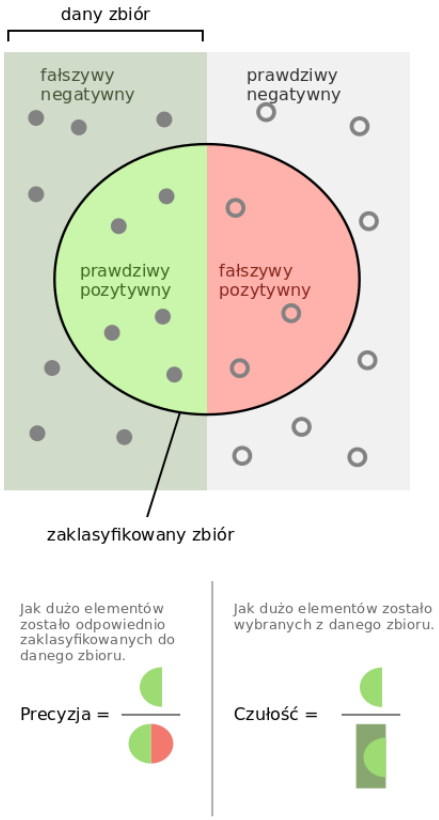
W celu ewaluacji algorytmu klasyfikacji zbiór wejściowy jest dzielony na dwa podzbiory zbiór treningowy oraz zbiór OOB (ang. *out-of-bag*). Na podstawie zbioru OOB można określić ile elementów zbioru treningowego zostało błędnie sklasyfikowanych. Lasy losowe nie wymagają wiedzy eksperckiej i są odporne 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 (13):

$$dokładność = \frac{Liczba\ poprawnych\ predykcji}{Całkowita\ liczba\ predykcji} \tag{13}$$

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*. Rys. 5. przedstawia ich graficzną zależność. Grafika została zaczerpnięta z [23] 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 (14).

$$czułość = \frac{prawdzie\ pozytywy}{prawdzie\ pozytywy + fałszywe\ negatywy} \tag{14}$$

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 (nieprawidłowe). W przypadku patologii głosu prawdziwymi pozytywami są prawidłowo określeni chorzy, a fałszywe negatywy są osobami które zostały zaklasyfikowane jako zdrowe, a faktycznie są chorzy. Czułość, to zdolność modelu do znalezienia wszystkich punktów w zbiorze danych, które są dla nas ważne.

Precyzja

Precyzja (15) jest definiowana jako liczba prawdziwych pozytywów podzielona przez liczbę prawdziwych pozytywów plus liczba fałszywych pozytywów.

$$precyzja = \frac{prawdziwe\ pozytywy}{prawdziwe\ pozytywy + fałszywe\ pozytywy} \tag{15}$$

Fałszywe pozytywy to przypadki, w których model błędnie określa jako pozytywne lub w naszym przykładzie osoby, które model uznaje za chorych, a które są w rzeczywistości zdrowe. Podczas, gdy czułość wyraża możliwość znalezienia wszystkich istotnych punktów w zestawie danych, precyzja wyraża proporcję punktów danych, które nasz model zaklasyfikowała poprawnie, a całkowitą ilość prognoz.

Wynik „F1”

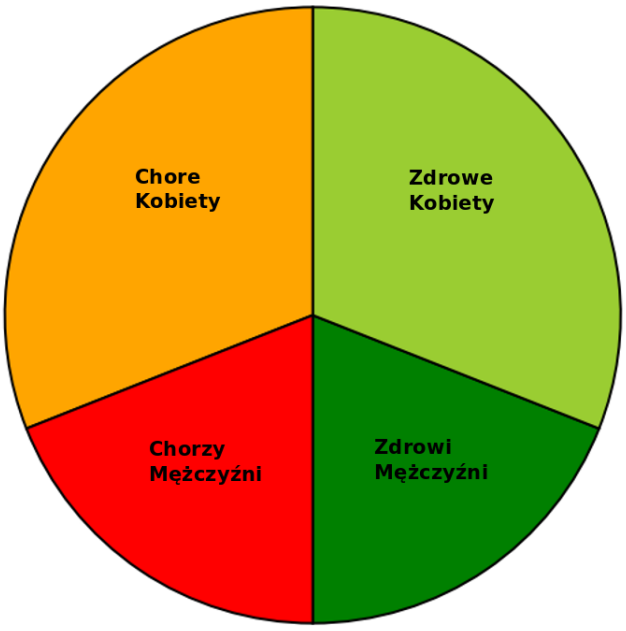
W przypadkach, w których chcemy znaleźć optymalną równowagę w precyzji i czułości, możemy połączyć te wskaźniki za pomocą tak zwanego wyniku F1. Wynik F1 jest średnią harmoniczną precyzji i czułości (16):

$$F1 = 2 * \frac{czułość * precyzja}{czułość + precyzja} \tag{16}$$

Jeśli chcemy znaleźć zbalansowany model klasyfikacji z optymalną równowagą pomiędzy czułością, a precyzją należy dążyć do maksymalizacji wyniku F1.

7. Wyniki badań

Ten etap badań ma na celu zaprezentowanie projektów systemów detekcji patologii 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 bez patologii głosu. Jako bazę danych wykorzystano zbiór nagrań głosowych pacjentów o nazwie *Saarbruecken Voice Database* (SVD). Szczegółowy opis bazy SVD znajduje się w Rozdziale 3. Wstępnie 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. Natomiast ocena jakościowa finalnej wersji systemu detekcji patologii zostanie oparta o wszystkie trzy głoski w normalnej tonacji. Aby skonstruować zbiór danych wybrałem 428 zdrowych i 428 chorych nagrań głosów kobiet, u mężczyzn liczby wyniosły 259 zdrowych i 259 chorych przypadków, Rys. 6.



Rys. 6. Relacja 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. Wektory cech charakterystycznych nagrań fonetycznych głosu, opracowałem na podstawie propozycji parametrów w Rozdziale 4. Umożliwiły one detekcję zniekształceń sygnału głosu. Ponieważ parametry różniły się zakresem zmienności oraz wartości, wektor parametrów poddałem standaryzacji opisanej wzorem (5). 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.

Aby mieć pewność do wyników klasyfikacji i mieć porównanie wybrałem z biblioteki *scikit-learn* [3] dwa różnie działające algorytmy: logistycznej regresji i metody lasów losowych opisanych w Rozdziale 5. Zbiór danych opisany poprzez wektor parametrów został podzielony w stosunku: 75% danych wykorzystano do treningu systemów uczących się, a pozostałe 25% danych posłużyło za test wytrenowanych algorytmów. Dodatkowo dla każdego z algorytmów zbiór treningowy podzieliłem na pięć części, aby przeprowadzić k-krotną walidację krzyżową (ang. *k-fold cross-validation*).

Ideą walidacji krzyżowej jest podział bazy danych losowo na 5 podobnych rozmiarowo podzbiorów. Jeden z podzbiorów ma za zadanie imitację zbioru treningowego, a reszta jest zbiorem, na którym zostaje wytrenowany algorytm klasyfikacji. Jako metryką oceny prognozy w procesie walidacji krzyżowej posłużyłem się dokładnością (13). Proces jest powtarzamy 5 razy, aż każdy podzbiorów znajdzie się w roli zbioru walidacyjnego. Walidacja krzyżowa została przeprowadzona osobno dla każdego z algorytmów i dla każdej płci z osobna.

Wykorzystanie k-krotnej walidacji pozwoliło na regulację *hiper-parametrów* każdego z algorytmów i oceny ich jakości bez użycia zbioru testowego. W trakcie projektowania algorytmów systemów uczących jest bardzo ważnym, aby zbiór testowy używać jak najrzadziej jest to możliwe, gdyż unikamy ryzyka dopasowania algorytmu do tych danych, co pozwoli uniknąć potencjalnego przetrenowania. Zadaniem badacza jest tak zaplanować eksperyment, aby dobrać jak najlepiej sprawdzające się *hiper-parametry*, ponieważ algorytm nie jest w stanie określić ich w trakcie treningu, a ich wartość ustawiana jest przed procesem nauczania.

W przypadku algorytmu logistycznej regresji będzie poszukiwana optymalna wartość współczynnika regularyzacji λ, który ma wpływa na wynik funkcji kosztu (10). Dla algorytmu losowych lasów poszukiwaną wartością będzie liczba drzew decyzyjnych, które będą uczestniczyć w procesie treningu i klasyfikacji. Liczba drzew użytych przy treningu i klasyfikacji jest kluczowym *hiper-parametrem* algorytmu loso-

wych drzew, zwiększenie ich ilości poprawia jakość prognozy, ale jednocześnie jest wysoce obciążające dla procesora.

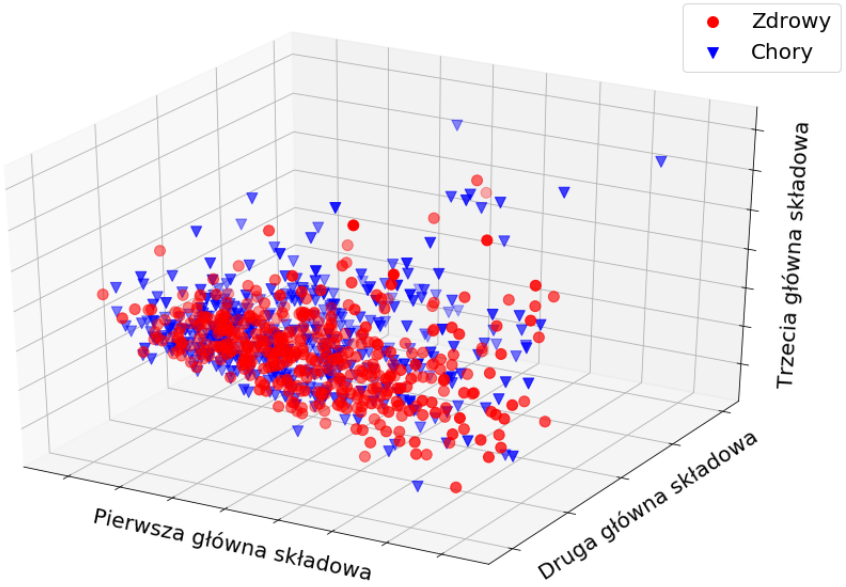
Kolejnym krokiem był trening algorytmów systemów uczących, których wyniki zostały przedstawione w tabelach. Ponadto jednym z atrybutów wyjściowych metody losowych lasów dostępnej w bibliotece *scikit-learn* jest wektor wag oparty na indeksie *Gini*. Każdy z parametrów posiada wagę przypisaną w czasie treningu informującą o jego ważności dla klasyfikacji. Wartości wag pozwalają wyznaczyć parametry o najwyższej wartości informacyjnej, w wyniku czego unikniemy cech o niskim wpływie na klasyfikację, poprzez ich wychwycenie i usunięcie. Wynik prognozy porównany z zestawem testowym 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. Następnie najlepiej sprawdzający się wektor parametrów w detekcji patologii i zweryfikowałem go na podstawie przedłużonych głosek /a/, /i/, /u/ w normalnej tonacji.

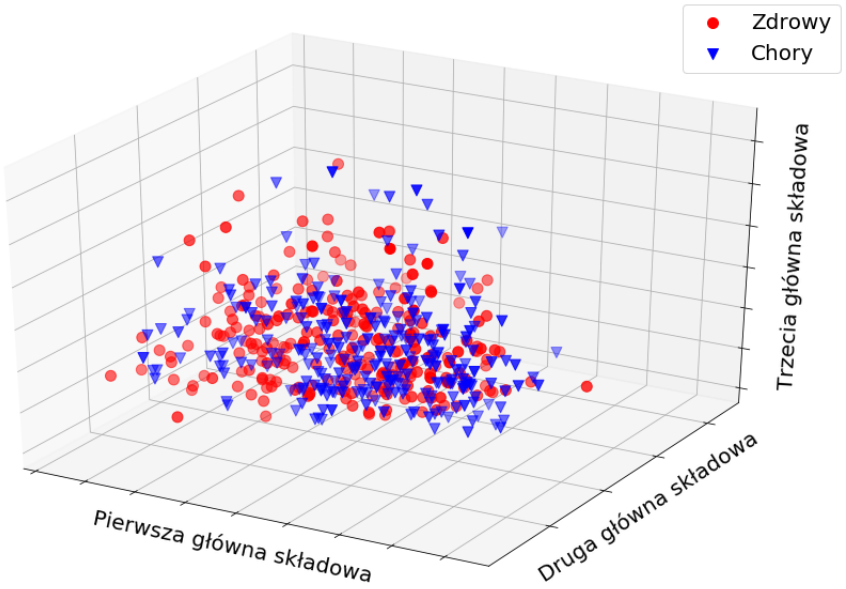
I próba

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 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 Rys. 7. i Rys. 8. przedstawiają analizę głównych składowych parametrów z rozróżnieniem na stan krtani pacjenta z podziałem na kobietę i mężczyznę. Rozmieszczenie punktów nie przejawia wyraźnej tendencji, która pozwoliłaby na ich klasyfikację opierając się na wizualnej reprezentacji. Nasuwa się wniosek, że wektory parametrów nie są wystarczające do przeprowadzenia detekcji deformacji mowy. Informacje dostarczone w postaci wartości charakterystycznych nagrań głosowych tworzą niepełny obraz krtani pacjenta. W wyniku czego system detekcji oparty na algorytmach systemów uczących się nie będzie w stanie poprawnie przeprowadzić klasyfikacji stanu pacjenta.

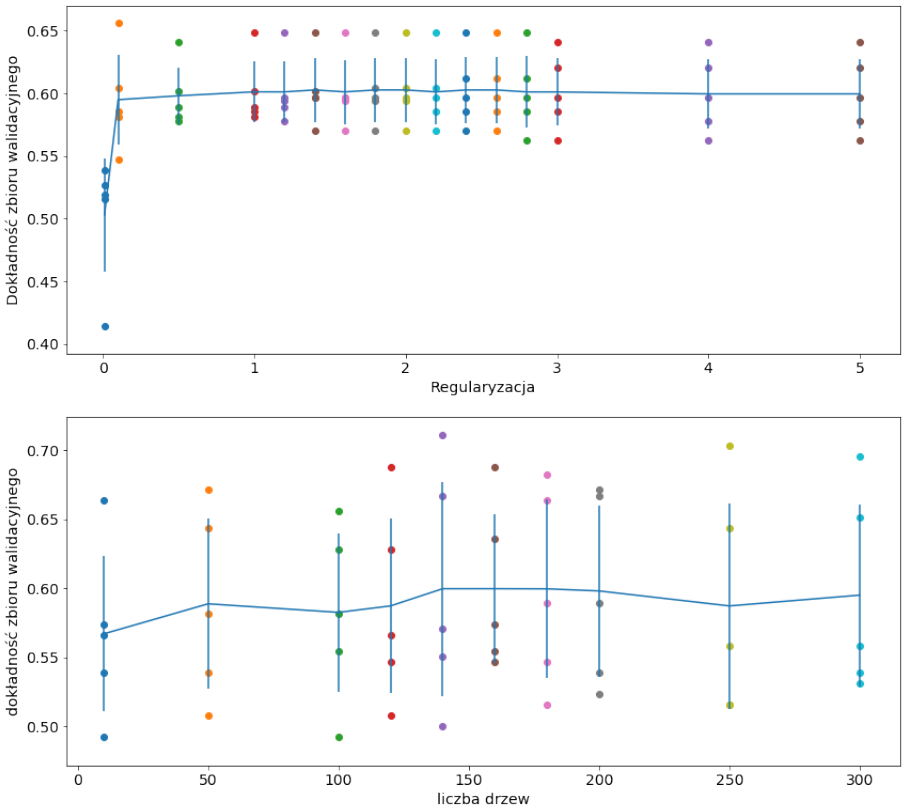


Rys. 7. Prezentacja 3 głównych głównych składowych samogłoski /a/ w normalnej tonacji dla kobiety z rozróżnieniem pomiędzy chorym, a zdrowym przypadkiem.

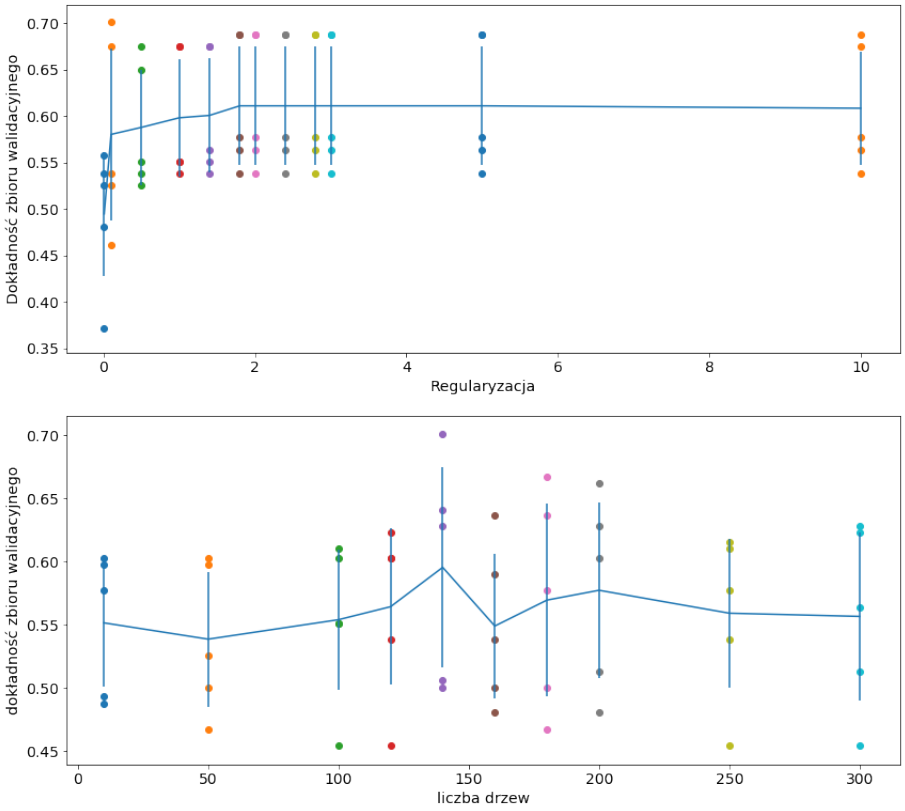


Rys. 8. Prezentacja 3 głównych składowych wektora parametrów samogłoski /a/ w normalnej tonacji dla mężczyzny z rozróżnieniem pomiędzy chorym, a zdrowym przypadkiem.

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*). Rysunki Rys. 9. i Rys. 10. przedstawiają poszukiwania współczynnika regularyzacji dla logistycznej regresji i liczby drzew decyzyjnych dla losowych lasów.



Rys. 9. Regulacja współczynnika regularyzacji dla logistycznej regresji i liczby drzew decyzyjnych dla losowych lasów w przypadku kobiet.



Rys. 10. Regulacja współczynnika regularyzacji dla logistycznej regresji i liczby drzew decyzyjnych dla losowych lasów w przypadku mężczyzn.

Jako kryterium wyboru świadczyła najwyższa dokładność zbioru walidacyjnego. Na ich podstawie wybrano następujące parametry:

- współczynnik regularyzacji, 2.6 dla kobiet i 2 dla mężczyzn,
- liczba drzew decyzyjnych, 160 dla kobiet i 140 dla mężczyzn.

Następnie wytrenowano dwa algorytmy z podanymi wyżej *hiper-parametrami*. Tabele Tab. 1. i Tab. 2. prezentują precyzję, czułość, wynik F1 i dokładność dwóch algorytmów: logistycznej regresji i losowego lasu dla kobiet i mężczyzn głoski /a/ w normalnej tonacji.

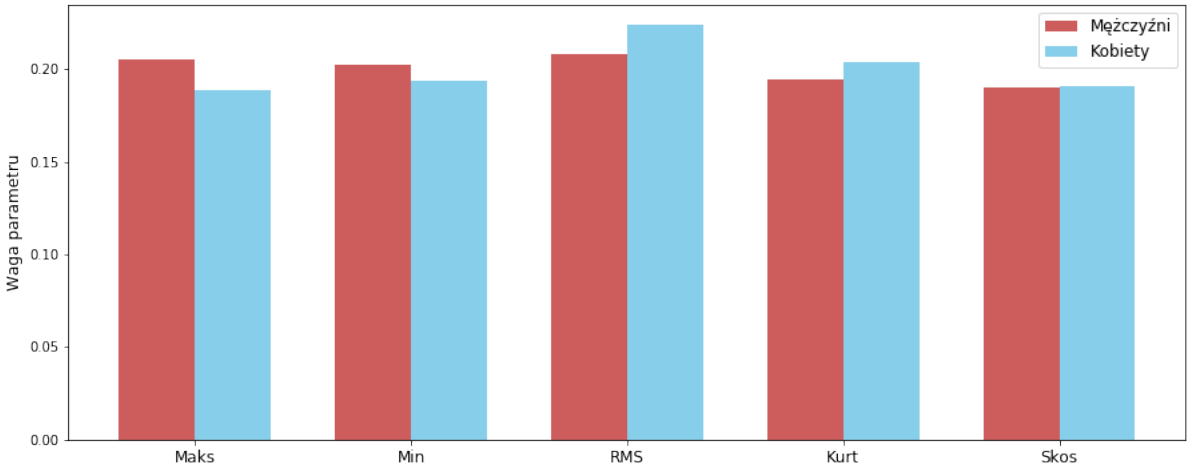
Tab. 1. Precyzja, czułość, wynik F1 i dokładność algorytmu logistycznej regresji dla kobiet i mężczyzn głoski /a/ w normalnej tonacji.

płeć	stan pacjenta	precyzja	czułość	wynik F1	dokładność
kobieta	zdrowy	0.59	0.68	0.63	0.60
	chory	0.62	0.53	0.57	
mężczyzna	zdrowy	0.55	0.61	0.58	0.54
	chory	0.53	0.46	0.49	

Tab. 2. 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	0.59	0.63	0.61	0.59
	chory	0.60	0.56	0.58	
mężczyzna	zdrowy	0.64	0.52	0.57	0.61
	chory	0.57	0.68	0.62	

Najmniejszą dokładność 54% osiągnął algorytm logistycznej regresji w klasyfikacji mężczyzn, wynik F1 wyniósł dla zdrowych 58%, 41% dla chorych. Największą dokładnością 61% wykazał się algorytm losowych lasów w klasyfikacji mężczyzn, wynik F1 wyniósł 57% dla zdrowych i 62% dla chorych. Systemy detekcji patologii nie są wystarczająco wytrenowane, dlatego należy zmodyfikować wektor parametrów wejściowych, aby poprawić ich skuteczność. Jednak wcześniej warto spojrzeć na wykresy wag każdego z parametrów, aby ocenić wartość informacyjną parametrów w czasie nauki algorytmu losowych drzew Rys. 11.



Rys. 11. Wagi dla każdego z parametrów przypisane przez algorytm losowych lasów w zależności od płci pacjenta.

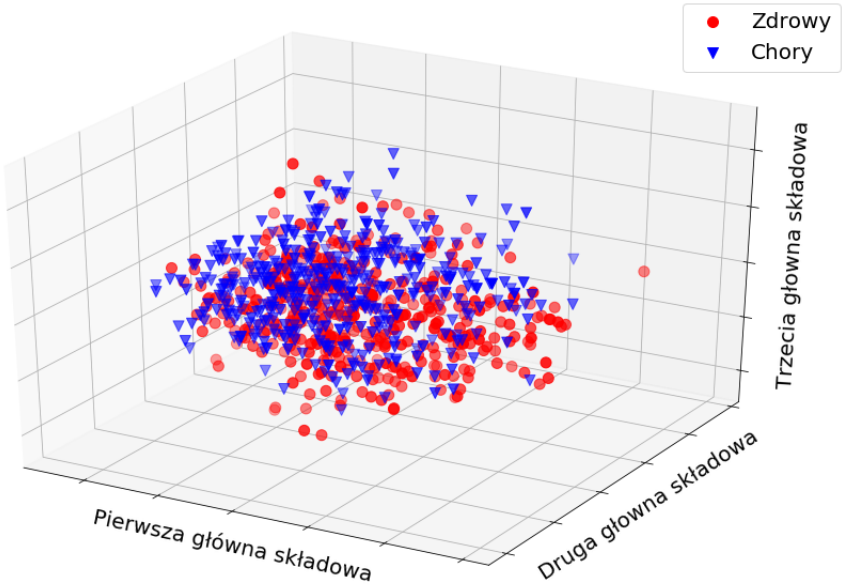
Brak wyraźnej granicy pomiędzy zdrowym, a chorym przypadkiem przekłada się na małą efektywność algorytmów. Potwierdzenie tej hipotezy pozwoliło wykazać, że algorytmy nie ulegają przetrenowaniu. Wektor parametrów zawiera dane, których rozkład nie wskazuje jednoznacznie na osobę zdrową lub chorą. Parametry otrzymane z analizy sygnału w dziedzinie czasu nie są wystarczającym źródłem do przeprowadzenia klasyfikacji, jednocześnie nie zważyłem wyraźnej różnicy informacyjnej pomiędzy cechami charakterystycznymi zbioru nagrań, dlatego wszystkie parametry zostaną użyte w następnej próbie. 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

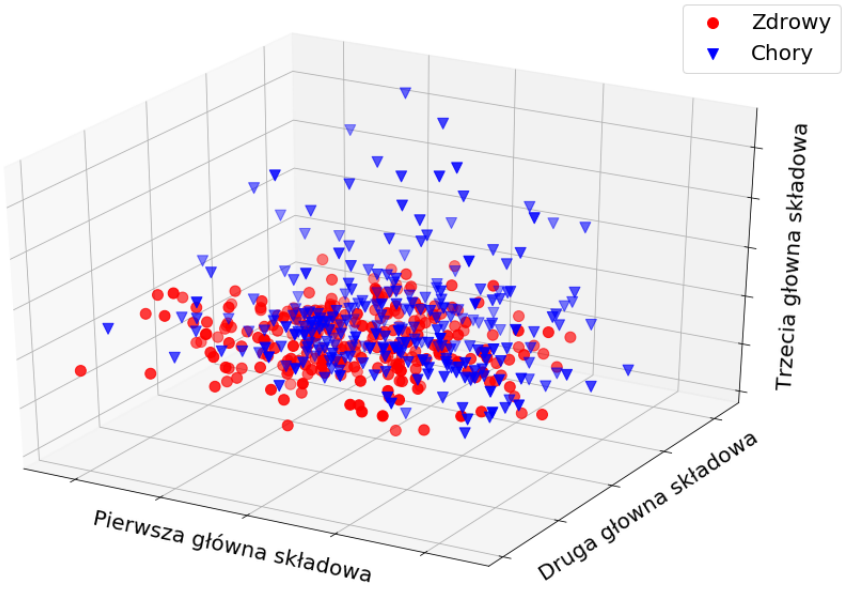
Poprzednia próba dowiodła, że analiza sygnału nagrań pacjentów w dziedzinie czasu nie jest wystarczająca dla poprawnej klasyfikacji, dlatego wektor uzupełnię o analizę sygnału w dziedzinie częstotliwościowej i cepstralnej. Nowy wektor parametrów będzie się składał z 17 parametrów: wartości minimalnej 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 [24]. Rysunki Rys. 12. i Rys. 13. przedstawiają analizę trzech głównych składowych wektora parametrów z rozróżnieniem na stan krtani pacjenta z podziałem na kobietę i mężczyznę.

Rysunki Rys. 14. i Rys. 15. przedstawiają poszukiwania współczynnika regularyzacji dla logistycznej regresji i liczby drzew decyzyjnych dla losowych lasów. Jako kryterium wyboru świadczyła najwyższa dokładność zbioru walidacyjnego. Na ich podstawie wybrano następujące parametry:

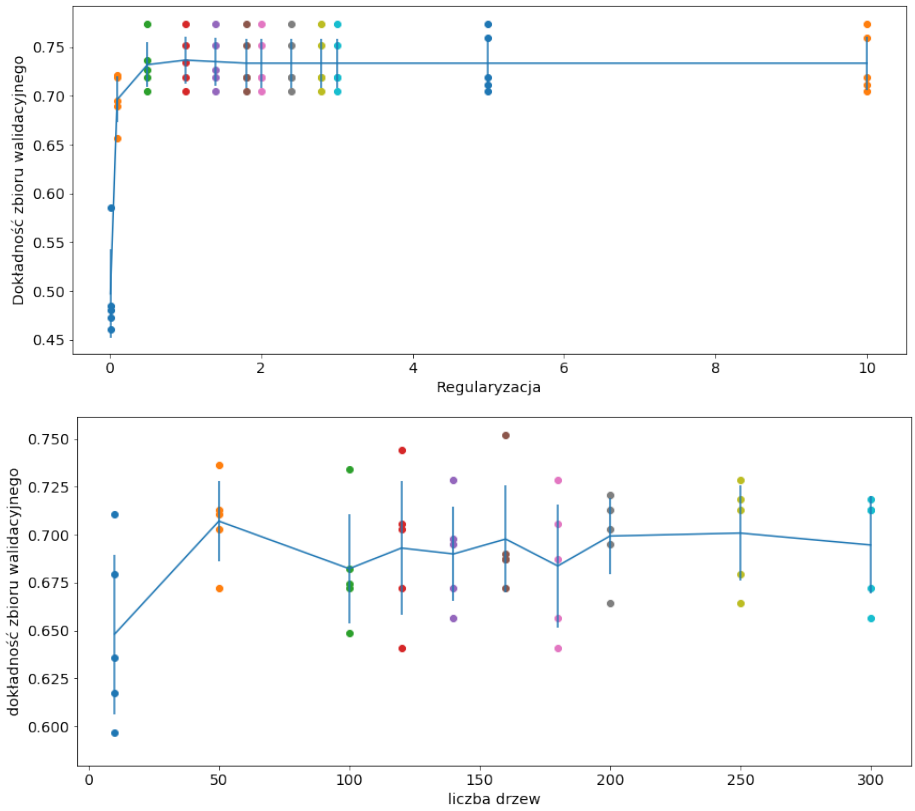
- współczynnik regularyzacji, 1 dla kobiet i 0.08 dla mężczyzn,
- liczba drzew decyzyjnych, 180 dla kobiet i 150 dla mężczyzn.



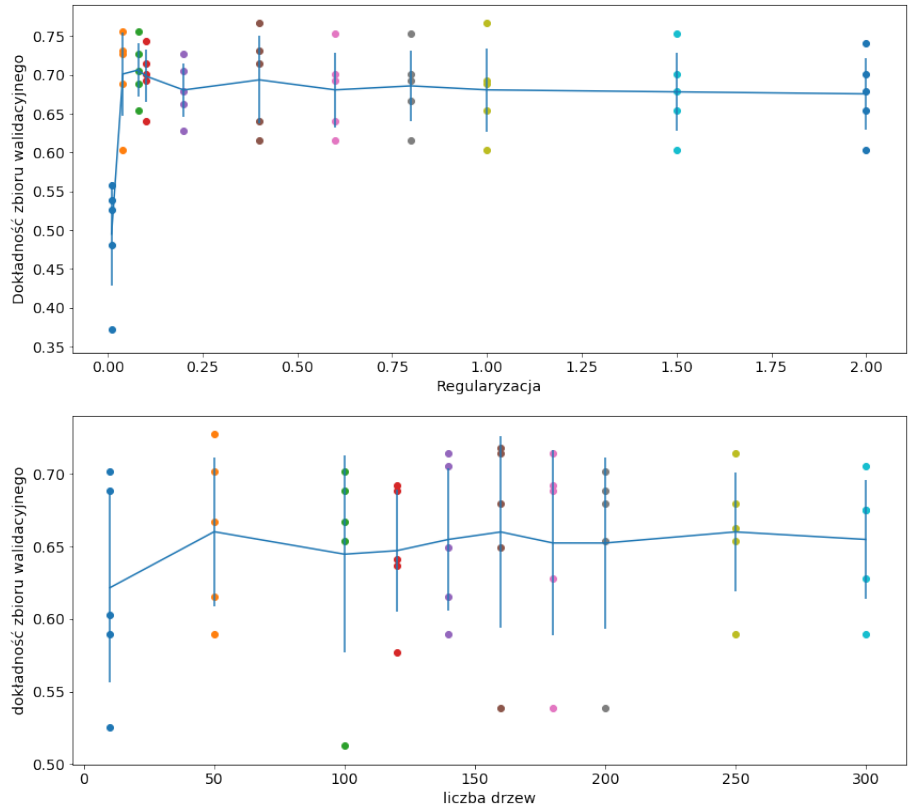
Rys. 12. Prezentacja 3 głównych głównych składowych samogłoski /a/ w normalnej tonacji dla kobiety z rozróżnieniem pomiędzy chorym, a zdrowym przypadkiem.



Rys. 13. Prezentacja 3 głównych składowych wektora parametrów samogłoski /a/ w normalnej tonacji dla mężczyzny z rozróżnieniem pomiędzy chorym, a zdrowym przypadkiem.



Rys. 14. Regulacja współczynnika regularyzacji dla logistycznej regresji i liczby drzew decyzyjnych dla losowych lasów w przypadku kobiet.



Rys. 15. Regulacja współczynnika regularyzacji dla logistycznej regresji i liczby drzew decyzyjnych dla losowych lasów w przypadku mężczyzn.

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 Tab. 3. i Tab. 4. prezentują precyzję, czułość, wynik F1 i dokładność tych dwóch algorytmów.

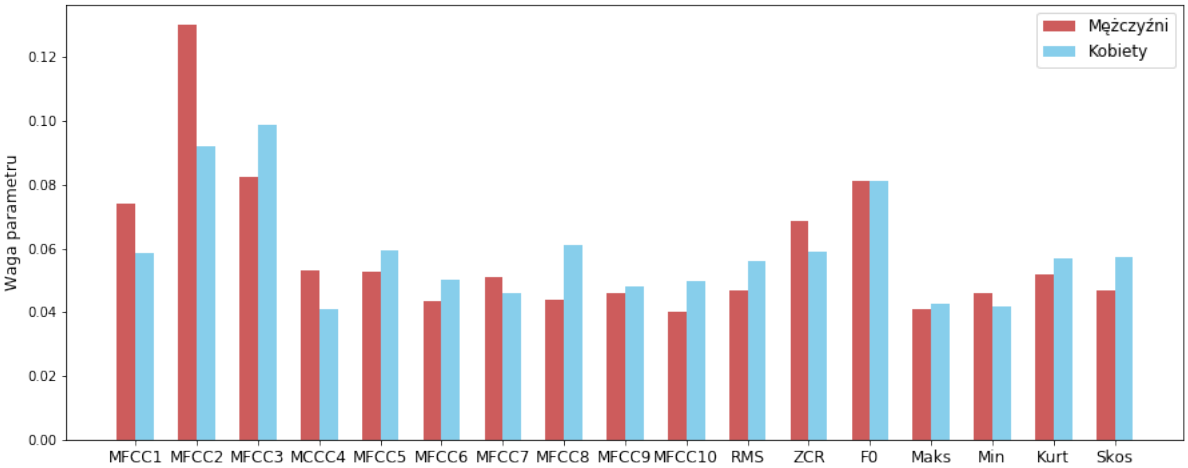
Tab. 3. Precyzja, czułość, wynik F1 i dokładność algorytmu logistycznej regresji dla kobiety i mężczyzny głoski /a/ w normalnej tonacji.

płeć	stan pacjenta	precyzja	czułość	wynik F1	dokładność
kobieta	zdrowy	0.75	0.77	0.76	0.75
	chory	0.75	0.73	0.74	
mężczyzna	zdrowy	0.72	0.64	0.68	0.68
	chory	0.66	0.73	0.69	

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	0.73	0.79	0.76	0.74
	chory	0.76	0.70	0.73	
mężczyzna	zdrowy	0.75	0.67	0.71	0.72
	chory	0.69	0.76	0.72	

W porównaniu z I próbą można zaobserwować wyraźną poprawę w klasyfikacji. Najmniejszą dokładność 72% osiągnął algorytm logistycznej regresji w klasyfikacji mężczyzn, wynik F1 wyniósł dla zdrowych 68%, 69% dla chorych. Największą dokładnością 75% wykazał się algorytm logistycznej regresji w klasyfikacji kobiet, wynik F1 wyniósł 76% dla zdrowych i 74% dla chorych. Rysunek Rys. 16. przedstawia wektor wag przypisanych dla każdego z parametrów w czasie treningu algorytmu losowych lasów. Wagi parametrów różnią się i zależą płci pacjenta.



Rys. 16. Wagi dla każdego z parametrów przypisane przez algorytm losowych lasów w zależności od płci pacjenta.

Najważniejsze parametry to MFCC2, MFCC3 i F_0. 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 zbędne.

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 dokładność 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.

Finalny system detekcji

Na podstawie powyższych badań wybrany został wektor parametrów o największej dokładność w wykrywaniu patologii. Po analizie uzyskanych wyników w każdej z prób, okazuje się, że najbardziej wydajny wektor składa się 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). Wybrany wektor zweryfikowany zostanie na tle trzech przedłużonych głosek /a/, /i/, /u/ w normalnej tonacji.

Dla każdego z algorytmów przeprowadzono regulację *hiper-parametrów*, aby uzyskać jak najlepsze wyniki. Regulację przeprowadzono dla każdej z głosek osobno i indywidualnie dla każdej płci. Tabele Tab. 5 i 6 przedstawiają optymalne wartości współczynnika regularyzacji i liczby losowych drzew wyznaczone na drodze krzyżowej walidacji.

Tab. 5. Regulacja współczynnika regularyzacji logistycznej regresji dla głosek /a/, /i/, /u/ w normalnej tonacji.

Głoska	Płeć	Współczynnik regularyzacji
/a/	kobieta	1
	mężczyzna	0.08
/i/	kobieta	0.08
	mężczyzna	0.04
/u/	kobieta	0.8
	mężczyzna	0.6

Tab. 6. Regulacja liczby drzew losowych lasów dla głosek /a/, /i/, /u/ w normalnej tonacji.

Głoska	Płeć	Liczba drzew
/a/	kobieta	180
	mężczyzna	150
/i/	kobieta	140
	mężczyzna	100
/u/	kobieta	140
	mężczyzna	140

Tabele Tab. 7 i Tab. 8 prezentują precyzję, czułość, wynik F1 i dokładność dwóch algorytmów losowych lasów i logistycznej regresji dla wszystkich trzech głosek.

Tab. 7. Precyzja, czułość, wynik F1 i dokładność algorytmu logistycznej regresji w wykryciu patologii dla kobiet i mężczyzny głosek /a/, /i/, /u/ w normalnej tonacji.

Głoska	Płeć	Precyzja	Czułość	Wynik F1	Dokładność
/a/	kobieta	0.75	0.77	0.76	0.75
	mężczyzna	0.69	0.77	0.72	0.72
/i/	kobieta	0.66	0.68	0.67	0.68
	mężczyzna	0.73	0.75	0.74	0.75
/u/	kobieta	0.64	0.69	0.66	0.65
	mężczyzna	0.77	0.73	0.75	0.75

Tab. 8. Precyzja, czułość, wynik F1 i dokładność algorytmu losowych lasów w wykryciu patologii dla kobiet i mężczyzny głosek /a/, /i/, /u/ w normalnej tonacji.

Głoska	Płeć	Precyzja	Czułość	Wynik F1	Dokładność
/a/	kobieta	0.73	0.79	0.76	0.74
	mężczyzna	0.66	0.73	0.69	0.68
/i/	kobieta	0.67	0.74	0.70	0.70
	mężczyzna	0.66	0.73	0.69	0.68
/u/	kobieta	0.63	0.66	0.65	0.65
	mężczyzna	0.73	0.66	0.69	0.70

Dokładność otrzymanych wyników mieści się w przedziale od 65% do 75%, każdy z algorytmów osiąga zbliżone wyniki w ocenie jakościowej.

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.

Praca przedstawia różne podejścia do projektowania i implementacji kompletnych systemów klasyfikacji. 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 przedłużonych samogłosek /a/, /i/, /u/ w normalnej tonacji. Zbiór danych zaczerpnięto z bazy *Saarbruecken Voice Database* (SVD), szerzej opisany w Rozdziale 3. Zaprojektowano model klasyfikacji binarnej, który segreguje pacjentów na zdrowych i chory.

Podczas badań zauważono niewielki odsetek prac, w których wybrano język *Python* do implementacji systemów detekcji patologii mowy. Dodatkowo wybrane środowisko pracy nie posiada dedykowanych narzędzi lub bibliotek do dokładnej analizy akustycznej sygnału audio pod kątem detekcji patologii mowy. 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. Badania przedstawiają możliwe rozwiązania wraz z ich oceną jakościową.

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, Rys.16.

Wszystkie algorytmy zaimplementowane podczas prac nad badaniami dostępne są w moim repozytorium na stronie GitHub [25].

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. Wszołek, "A vibroacoustic model of selected human larynx diseases," *International Journal of Occupational Safety and Ergonomics*, vol. 13, no. 4, s. 367–379, 2007.

[3] scikit-learn. Machine Learning in Python. [Online]. Available: https://scikit-learn.org/stable/index.html

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[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] C. T. Herbst, W. T. S. Fitch, J. G. Švec, "Electroglottographic wavegrams: a technique for visualizing vocal fold dynamics noninvasively," *The Journal of the Acoustical Society of America*, vol. 128, no. 5, s. 3070–3078, 2010.

[13] File extension .wav. Wikiepdia. [Online]. Available: https://en.wikipedia.org/wiki/WAV

[14] D. Hemmerling, *The use of the speech signal as a source of diagnostic, control and forecasting information in selected medical problems related to otorhinolaryngology*, 2018.

[15] B. Wilk, "Wyznaczanie wartości chwilowej częstotliwości podstawowej tonu krtaniowego za pomocą analizy falkowej sygnału mowy," *Przegląd Elektrotechniczny*, vol. 91, s. 305–308, 2015.

[16] T. P. Zieliński, *Cyfrowe przetwarzanie sygnałów: od teorii do zastosowań*. Wydawnictwa Komunikacji i Łączności, 2007.

[17] A. Izworski, R. Tadeusiewicz, W. Wszolek, "Artificial intelligence methods in diagnostics of the pathological speech signals," in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. Springer, 2004, s. 740–748.

[18] A. M. Noll, "Short-time spectrum and cepstrum techniques for vocal-pitch detection," *The Journal of the Acoustical Society of America*, vol. 36, no. 2, s. 296–302, 1964.

[19] S. N. Awan, A. Giovinco, J. Owens, "Effects of vocal intensity and vowel type on cepstral analysis of voice," *Journal of voice*, vol. 26, no. 5, s. 670–e15, 2012.

[20] V. S. Chakravarthi, Y. J. M. Shirur, P. Rekha, *Proceedings of International Conference on VLSI, Communication, Advanced Devices, Signals & Systems and Networking (VCASAN-2013)*. Springer Science & Business Media, 2013, vol. 258.

[21] 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.

[22] B. Leo, J. H. Friedman, R. A. Olshen, C. J. Stone, "Classification and regression trees," *Wadsworth International Group*, 1984.

[23] Walber. (2014) Precision and recall. [Online]. Available: https://commons.wikimedia.org/wiki/File:Precisionrecall.svg

[24] Librosa. Librosa Development Team. [Online]. Available: http://librosa.github.io/librosa/index.html#librosa

[25] B. Tynski. Voice pathology detection in python. [Online]. Available: https://github.com/tynski/voicePathology