**Keywords:**

heartbeat classification, arrhythmia, ECG signal, deep learning, spectrogram

**Abstract.**

Stworzone rozwiązanie pozwala na analizę sygnału elektrokardiograficznego. Praca opisuje proces formułowania klasyfikatora na podstawie informacji zawartych w bazie arytmii MIT-BIH. Uzyskane wyniki klasyfikacji są na wysokim poziomie i mogą konkurować ze znanymi rozwiązaniami z ostatnich lat.

Szczegółowo opisano podjęte kroki prowadzące do uzyskania przewidywań odnośnie typu arytmii dla każdego uderzenia serca. Zaawansowane zagadnienia, które sprawiały trudność podczas implementacji przedstawiono w precyzyjny sposób. Dokonano przeglądu rozwiązań, by przybliżyć kilka najczęstszych podejść do rozpatrywanego problemu.

Projekt wykorzystuje publicznie dostępną bazę danych. Znajdują się w niej sygnały elektrokardiograficzne z dwóch czujników. Oba wykorzystano, co nie jest częstym zjawiskiem dla algorytmów tego typu. Klasyfikator w procesie uczenia wykorzystuje tylko informacje o dużej pewności. Dane bazują na adnotacjach ekspertów, a na przestrzeni lat znalezione błędy zostały poprawione.

Typy uderzeń serca podzielono na specjalne grupy według standardu „Association for the Advancement of Medical Instrumentation” (AAMI). Zaleca on pogrupowanie typów na pięć poszczególnych grup według fizjologicznego pochodzenia. Liczność rzadkich typów wykorzystanego źródła danych była niewystarczająca. Dla jednej grupy zastosowano metody modyfikujące, które pozwoliły zwiększyć ilość danych zbiorów treningowych. Przełożyło się to pozytywnie na wyniki.

W skład rozwiązania wchodzi ekstrakcja cech. Wykorzystano jedną próbkę danych do zapisania informacji na temat dwóch sygnałów. Jest to niespotykany sposób przetwarzania, który pozwolił na osiągnięcie dobrych wyników.

Głównym modułem klasyfikatora jest głęboka sieć neuronowa. Wybrano rozwiązanie przynoszące najlepsze efekty i przy użyciu narzędzi wspomagających automatyczny dobór hiperparametrów odnaleziono optymalne wartości.

Przeprowadzono analizę wyników klasyfikacji i porównano efekty z pracami o podobnej tematyce. Wykorzystano macierz błędów do znalezienia zależności między predykcją poszczególnych klas. W diagnostyce sygnałów EKG najważniejszym zadaniem jest rozdział grupy pobudzeń nadkomorowych i pobudzeń komorowych. W pracy udało uzyskać się liczbę tych błędów na bardzo niskim poziomie 0,29 % wszystkich próbek tych grup. Na koniec opisano kilka propozycji co do rozbudowy rozwiązania.

**Introduction**

Choroby układy krwionośnego to nadal jedna z najczęstszych przyczyn śmierci. W wielu schorzeniach, w tym także kardiologicznych, decyzja o metodach ratowania życia i sposobach terapii zależ od szybkiego i prawidłowego rozpoznania jednostki chorobowej [4]. Rozpoznanie jest dokonywane przez lekarza na podstawie wywiadu z pacjentem oraz wyników przeprowadzonych badań. Wiele symptomów obejmuje całą gamę chorób, postawienie diagnozy z dużym prawdopodobieństwem nie jest prostym zadaniem. Lekarze z biegiem czasu będą używać w większym stopniu narzędzi, które w czasie rzeczywistym pomagają podjąć decyzję. Medycyna posiada już wiele różnych rozwiązań w większości bazujących na metodach sztucznej inteligencji. Algorytmy często wykorzystują sztuczne sieci neuronowe, sieci Bayesa, algorytmy genetyczne, drzewa decyzyjne i logikę rozmytą [4].

Gdy zachodzi potrzeba analizy krótkiego zapisu elektrokardiograficznego (EKG) doświadczony specjalista nie będzie miał z tym problemu. Badania przeprowadzane metodą Holtera mogą pochodzić z całego dnia pomiaru. Prawidłowa częstotliwość rytmu serca może być w przedziale od 60 do nawet 90 uderzeń w chwili spoczynkowej. Takie liczby dają niewyobrażalnie duży zbiór do przeanalizowania w ciągu całego dnia. Dlatego metody automatycznej klasyfikacji są potrzebne. Jako wynik analizy lekarz otrzymuje wybrane fragmenty sygnału oraz proponowane rodzaje patologii [5].

Rejestracja sygnałów medycznych w sposób ciągły to istotna część diagnostyki. Monitorowanie i nadzór bieżącego stanu pacjenta pozwala dostarczyć szereg informacji. Sygnał powstały z czujników podczas zabiegu diagnostycznego jest wykorzystywany w medycynie w celu rozpoznawania chorób serca. Pomiar EKG jest to rejestracja elektrycznej czynności mięśnia sercowego z powierzchni klatki piersiowej w postaci różnicy potencjałów (napięć) pomiędzy elektrodami, co graficznie odczytujemy w formie krzywej elektrokardiograficznej, na specjalnym papierze milimetrowym bądź na ekranie monitora. Warto zwrócić uwagę na sektor urządzeń mobilnych monitorujących parametry życiowe, powstają również rejestratory EKG jako moduły rozszerzające dowolną platformę mobilną, jak telefon czy tablet [6].

(cel i zakres pracy)

Głównym celem było stworzenie rozwiązania pozwalającego analizować sygnał elektrokardiograficzny z wysokim wynikiem poprawności klasyfikacji. Poszczególne kroki przetwarzania prowadzą końcowo do uzyskania przewidywań dotyczących każdego uderzenia serca. Zadaniem algorytmu jest wskazanie typu uderzenia serca z dużym poziomem pewności co do przewidywania.

Początkowo sygnał zostanie zdekodowany i zapisany w formie bardziej odpowiedniej dla dalszego przetwarzania. Faza zawiera również kroki mające na celu odrzucenie informacji mniej istotnych z punktu widzenia dalszej obróbki.

Następnym etapem jest ekstrakcja cech. Wyszczególnienie właściwości różniących określone typy wpłynie znacząco na wyniki. Ten proces wpływa również na informacje nie przydatne w przebiegu, dokonując ich redukcji.

Najważniejszym krokiem jest budowa klasyfikatora. Bazuje na sieci neuronowej o dużej ilości warstw ukrytych.

Dokonano oceny dokonując analizy wyników i ich porównanie z wynikami pochodzącymi z prac o podobnej tematyce.

**Methodology**.

Dane zostały wstępnie przetworzone by uzyskać lepsze efekty i wyniki. Czynności mające na celu uwydatnienie zależności znajdujących się w przetwarzanych informacjach nazywa się ekstrakcją cech.

Żeby dokonać poprawnej klasyfikacji typów uderzeń serca należy przejść kilka etapów analizy. Zapis sygnału posiadający zakłócenia może przejść proces filtracji w celu pozbycia się szumów i artefaktów. Następnym krokiem jest detekcja załamków sygnału EKG. Najbardziej dominujący w przebiegu czasowym jest zespół QRS. Niestety są typy uderzeń, które nie charakteryzują się taką budową. Załamki P oraz T mogą być szybkozmienne w badanym sygnale. Mogą się również zdarzyć składowe pochodzące z aparatury pomiarowej lub też stymulatora serca badanego pacjenta. Wpływa to znacząco na wyniki, gdy używane są dość proste algorytmy. Określenie dobrego położenia załamka QRS jest bardzo ważne podczas oddzielania istotnych przedziałów czasowych.

Istnieje dużo bibliotek odznaczających się małym błędem, jeśli chodzi o szukanie odpowiednich części w sygnale EKG. Do trenowania stworzonego modelu zostały użyte jedynie typy opisane przy pomocy bazy MIT-BIH. Taką metodologię wykorzystano również w [13], [20], [21] i [22]. Pozwala to uzyskać jak najmniejszy błąd wyniku i oceny modelu sieci neuronowej.

(trochę o bazie danych)

Duży przedział czasu w jakim ona istniała i jej dostępność sprawiły, że była często wykorzystywana przez badaczy. Zapisano ją używając formatu MIT. Zawiera 48 półgodzinnych fragmentów dwukanałowych ambulatoryjnych zapisów EKG. Zostały one pozyskane od 47 osób badanych w „Beth Israel Hospital Arrhythmia Laboratory”. Dwadzieścia trzy zapisy wybrano metodą losową z zestawu czterech tysięcy dostępnych zapisów EKG, pobranych od mieszanej populacji pacjentów hospitalizowanych (około 60%) i ambulatoryjnych szpitala w Bostonie. Pozostałe dwadzieścia pięć nagrań wybrano używając tego samego zestawu, ale z uwzględnieniem rzadszych i klinicznie istotnych zaburzeń rytmu, które nie byłyby dobrze reprezentowane używając metod losowych. W tych 25 rekordach znajdują się również przypadki złożonych arytmii komorowych, połączeniowych i nadkomorowych oraz zaburzeń przewodzenia. Takie zapisy ze względu na cechy rytmu, zmienności morfologii zespołu QRS lub jakości sygnału będą stanowić znaczną trudność dla detektorów arytmii. Baza danych była udoskonalana na przestrzeni lat. Zawiera zapisy badań EKG pacjentów borykających się z problemem arytmii serca.

(grupowanie uderzeń)

Liczność typów uderzeń serca przetwarzanej baza danych jest bardzo duża. Choć informacje w niej zawarte są dobrane w taki sposób by uwzględnić pewne specyficzne i rzadko spotykane schorzenia to liczba wystąpień kilku typów jest niewielka. Nie wystarczająca dla uzyskania dobrych wyników dla większości algorytmów. Wybrany został sposób grupowania który pozwoli osiągnąć dobre wyniki i podzielić typy uderzeń serca używając rozsądnych zależności. W tabeli 4.1 zestawione zostały wszystkie typy uderzeń serca wraz z ich ilością po preprocesingu. Liczba może się nieznacznie różnić ze względu na wykluczenie przedziałów, które występowały na początku lub końcu sygnału i nie posiadały odpowiedniej długości do przetworzenia w dalszych etapach. W tabeli znajdują się również akronimy używane w literaturze medycznej.

Tabela 4.1. Zestawienie klas arytmii oraz ich liczności.

|  |  |  |  |
| --- | --- | --- | --- |
| Typ uderzenia serca | Akronim | Oznaczenie | Ilość |
| Uderzenie normalne | NORMAL | N | 74984 |
| Blok lewej odnogi pęczka Hisa | LBBB | L | 8069 |
| Blok prawej odnogi pęczka Hisa | RBBB | R | 7250 |
| Zastępcze pobudzenie przedsionkowe | AESC | e | 16 |
| Zastępcze pobudzenie węzłowe | NESC | j | 229 |
| Przedwczesne pobudzenie przedsionkowe | APC | A | 2544 |
| Zablokowane przedwczesne pobudzenie przedsionkowe | ABERR | a | 150 |
| Przedwczesne pobudzenie węzłowe | NPC | J | 83 |
| Przedwczesne lub ektopowe pobudzenie nadkomorowe | SVPCs/SVE | S | 2 |
| Przedwczesne pobudzenie komorowe | PVC | V | 7128 |
| Zastępcze pobudzenie komorowe | VESC | E | 106 |
| Trzepotanie komór | FLWAV | ! | (472) |
| Fuzja pobudzenia komorowego i normalnego | FUSION | F | 802 |
| Pobudzenie wystymulowane | PACE | / | 7020 |
| Fuzja uderzenia wystymulowanego i normalnego | PFUS | f | 982 |
| Niesklasyfikowane pobudzenie | UNKNOWN | Q | 33 |

Warto wspomnieć o pewnej nieprawidłowości, opisy na oficjalnej stronie [23] z której pochodzi baza danych *MIT-BIH* nie zgadzają się z oznaczeniami istniejącymi w plikach przebiegów. Przeanalizowano liczby, sygnatury i wygląd poszczególnych typów by odpowiednio dobrać i zmienić oznaczenia. Przywoływane nazwy są spójne z przedstawionymi w plikach i będą używane w tej pracy. Typ o oznaczeniu „!” nie został w ogóle uwzględniony, głównie ze względu na brak pewności, gdzie powinien on przynależeć. W pracach [24], [16], [10] i [18] nie został on uwzględniony. Prace [6] i [13] zgrupowały go do grupy uderzeń komorowych. Jest jeszcze jeden powód wpływający na podjętą decyzję. Typ trzepotania komorowego występuje tylko w jednym zapisie o nazwie „207”. Oznacza to, że dotyczy on jednego pacjenta.

Należy zwrócić uwagę na dużą dominację uderzeń normalnych. Następne w kolejności są bloki odnóg pęczka Hisa, przedwczesne pobudzenie komorowe i pobudzenie wystymulowane. Choć ma to pewne odzwierciedlenie w prawdopodobieństwie wystąpienia to jednak taki rozkład może wpływać na ostateczne wyniki klasyfikatora.

W pracy został wykorzystany standard stworzony przez organizację „Association for the Advancement of Medical Instrumentation” (*AAMI*). Wspiera ona rozwój bezpiecznych i efektywnych technologii medycznych. Norma *AAMI* „EC38:2007” opisuje testowanie i raportowanie wyników pracy serca oraz algorytmy pomiaru rytmu i przebiegów uderzeń. Zaleca pogrupowanie typów na pięć szczególnych grup. Zgodnie z normą, całkowita baza danych *MIT-BIH* powinna być podzielona ze względu na ich fizjologiczne pochodzenie.

Warto wspomnieć, że wiele prac, nie jest w pełni zgodnych z normą *AAMI*, ponieważ klasy *NESC* i *AESC* są przydzielone do grupy "N" zamiast "S" [24]. Do tych prac należą: [13], [10] i [18]. Nieprawidłowe etykietowanie wpływa na 245 uderzeń w bazie danych, stanowi to 8% wszystkich uderzeń nadkomorowych. Dokładne grupowanie można zobaczyć w tabeli 4.2.

Tabela 4.2. Pogrupowanie typów uderzeń serca według standardu *AAMI*.

|  |  |  |  |
| --- | --- | --- | --- |
| Grupa uderzeń serca | Grupa standardu AAMI | Typ bazy MIT-BIH | Ilość |
| Uderzenie normalne | N | NORMAL, LBBB, RBBB | 90303 |
| Uderzenie nadkomorowe | S | APC, NPC, ABERR, SVPCs/SVE, NESC, AESC | 3024 |
| Uderzenie komorowe | V | PVC, VESC | 7234 |
| Fuzje uderzeń | F | FUSION | 802 |
| Uderzenie niesklasyfikowane | Q | PACE, PFUS, UNKNOWN | 8035 |

Grupa uderzeń normalnych jest największa. To bardzo duża dysproporcja, zwłaszcza że nie są to typy uderzeń serca na których powinien się skupić specjalista podczas analizowania wyników.

(przetwarzanie danych)

W ramach prezentowanej pracy stworzono przy użyciu języka „Python” metody wczytywania zapisów bazy MIT-BIH. Każdy z dwóch sygnałów jest wczytywany przy użyciu biblioteki WFDB („Wave Form Data Base”) [25] oraz zapisywane w tablicy biblioteki „NumPy”. Przy pomocy wydajnych narzędzi czas wczytania, przetworzenia i zapisania przydatnych informacji jest niewielki. Nazwa bazy danych wskazuje format jej zapisu. Struktura zapisu MIT opisuje przechowywanie danych. Trzy podstawowe pliki są odpowiedzialne za magazynowanie nagrań sygnałów EKG. Dokument nagłówkowy zawiera tekst w formacie ASCII („American Standard Code for Information Interchange”), dotyczący tytułu nagrania, liczby oraz nazwy kanałów sygnału. Posiada on również informacje o częstotliwości próbkowania sygnału, długości zapisu, formacie binarnym sygnału. Może zawierać również dane o wieku i płci badanej osoby, dołączone są też informacje o diagnozie lekarskiej i przyjmowanych lekach przez pacjenta. Następny plik jest oryginalnym zapisem sygnału w postaci binarnej, w tym przypadku są to dwa przebiegi. Ostatnim dokumentem formatu MIT jest plik z adnotacjami czasowymi. Każdy zapis jest kodowany w 16 bitach. Sześć najważniejszych odpowiada za zakodowanie typu morfologii uderzenia lub zmiany rytmu serca. Pozostałe bity wyrażają odległość od poprzedniej adnotacji.

Każdy z pojedynczych plików jest nazwany liczbą pochodzącą od nazwy badania. Adnotacje pozwalają na umiejscowienie załamka R w sygnale. Taka wskazówka pozwala na wycięcie istotnych przedziałów i tylko one są dalej przetwarzane. Zasięg obejmuje 260 punktów sygnału z próbkowaniem 360 Hz. Punkt adnotacji, czyli pozycja załamka R została przesunięta o odległość równą 40% całego zakresu licząc od pierwszego punktu wektora. Ta wartość została wybrana ze względu na szczególnych charakter części sygnału charakteryzującej uderzenie serca. Część istotna przed załamkiem R jest krótsza niż połowa długości wyciętego wektora. Na tym etapie tracona jest również informacja o przynależności uderzeń do konkretnych pacjentów. Ten kroki nie są wymagające, jeśli chodzi o zasoby komputera i przebiegają bardzo szybko.

Wykorzystanie dwóch dostępnych sygnałów zdarza się tylko w nielicznych pracach jedną z nich jest [16], takie podejście pozwala na uzyskanie lepszych wyników. Ze względu na mnogość rozwiązań detekcji zespołów QRS w sygnale EKG nie jest tu badana ich dokładność. Początkowe testy wykazały, że poprawność ogólnodostępnych rozwiązań jest bardzo dobra i obarczona tylko niewielkim błędem. Można znaleźć wiele udostępnionych bibliotek [26], które bazują na algorytmach z różnym poziomem poprawności.

Zapis sygnału EKG jest obarczony szeregiem zakłóceń. Na przebieg czasowy może wpływać sposób zasilania osprzętu pomiarowego, charakterystyka miejsca podłączenia elektrody oraz ruchy pacjenta podczas badania. Częstym problemem jest wahanie poziomu linii izoelektrycznej, którego źródłem mogą być ruchy oddechowe klatki piersiowej [27].

(modyfikacja fuzji uderzeń)

W rzeczywistych zastosowaniach często wykorzystywana jest augmentacja danych. Głębokie sieci neuronowe potrzebują dużej ilości informacji, żeby nauczyć się określnych cech danej klasy. Powielenie przez modyfikację pozwala w prosty sposób zwiększyć liczbę zasobności zbioru. Należy pamiętać by dane po modyfikacji nie utraciły istotnych informacji. Grupa fuzji uderzeń posiada bardzo mało przykładów. W większości prac osiąga najgorszy wynik podczas klasyfikacji. Ze względu na to, że w tym zbiorze znajduje się tylko jeden typ uderzeń serca, łatwiej go przetworzyć i zweryfikować, jak modyfikacja wpłynie na wynik.

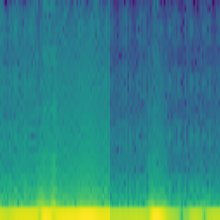
Modyfikacja została dokonana bazując na wcześniej utworzonych plikach z wektorami reprezentującymi przedziały uderzeń serca. Dane w ilość 80% całej grupy zostały powielone czterokrotnie. Proces rozpoczęto od odrzucenia danych idących na poczet zbioru testowego. Pozostałe informacje obsłużono przy użyciu funkcji reprezentującej pierwszą metodę modyfikacji. Jej zadaniem jest wykonanie mnożenia informacji pochodzących z dwóch sygnałów. Oznacza to użycie dwóch wektorów dotyczących tej samej adnotacji. Liczba mnożnika jest losowana przed każdym wyborem następnej jednostki danych. Wartość mieści się w przedziale od 0.95 do 1.05.

Modyfikacje tu opisane pozwalają stworzyć pewne odmienności nie burząc w znacznym stopniu charakterystyki sygnału.

(tworzenie spektrogramu)

Jednym z często wykorzystywanych kroków podczas preprocesingu sygnału *EKG* jest tworzenie spektrogramu. To analiza czasowa widma, pokazuje w dziedzinie czasu i częstotliwości jaki jest rozkład energii badanego przebiegu czasowego. Wyznacza się go, dzieląc sygnał na krótsze okresy, dla których obliczane są amplitudy składowych harmonicznych [28]. Zastosowanie tej metody w następnych etapach przetwarzania sygnału pozwala na ominięcie wielu kroków które by występowały w fazie wstępnej. Spektrogram jest tworzony przy użyciu kroczącej transformaty Fouriera, obrazuje zmiany częstotliwościowe. Oznacza to, że znaczna część błędów jest tracona podczas tej ekstrakcji cech.

Proces tworzenia spektrogramu dotyczy obu sygnałów. Wynikowy obraz z przestrzenią kolorów RGB jest wytworem składającym się z dwóch przetworzeń. Na wyjściu powstaje obraz złączenia dwóch spektrogramów. Stworzono go w taki sposób, że lewa strona przestawia wynik kroczącej transformaty Fouriera pierwszego sygnału, a prawa drugiego. Tworząc razem zdjęcie o wymiarach 220 pikseli wysokości i tyle samo szerokości.



Rysunek 4.3. Zdjęcie końcowe składające się z dwóch spektrogramów, pochodzi z grupy uderzeń komorowych.

(podział danych)

W prezentowanej pracy wykorzystano metodę uczenia nadzorowanego wszystkich typów uderzeń pochodzących z bazy danych *MIT-BIH*. Klasyfikacja jest dokonywana dla prawie każdego pobudzenia. Zostały odrzucone początkowe i końcowe przedziały każdego zapisu. Powodem była niewystarczająca długość sygnału. By wykorzystać te części należałoby przeprowadzić dodatkowe przetwarzanie by nie wprowadzać błędów do zbioru.

Podział danych nastąpił przy użyciu biblioteki „Split Folders”. Przekazano 60 % na dane testowe i 20 % na walidacyjne. Zbiór testowy również zawierał 20 % zbioru początkowego. Po tym procesie zostały dodane informacje zmodyfikowane. Zbiór treningowy reprezentuje wszystkie grupy uderzeń serca, ich dokładna liczność została przedstawiona w tabeli 4.3.

Tabela 4.3. Liczność klas w zbiorze treningowym, walidacyjnym i testowym.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Zbiór danych | Liczba jednostek uderzeń serca danej grupy | | | | |
| Normalne | Nadkomorowe | Komorowe | Fuzje | Niesklasyfikowane |
| Treningowy | 54181 | 1814 | 4340 | 2472 | 4821 |
| Walidacyjny | 18060 | 604 | 1446 | 898 | 1607 |
| Testowy | 18062 | 606 | 1448 | 160 | 1607 |

Testowy zbiór danych reprezentuje rozkład poszczególnych grup całej bazy. Pozostałe dwa zbiory posiadają dodatkowe modyfikowane jednostki pochodzące z grupy fuzji uderzeń. Proces modyfikacji został przedstawiony w poprzednich rozdziałach.

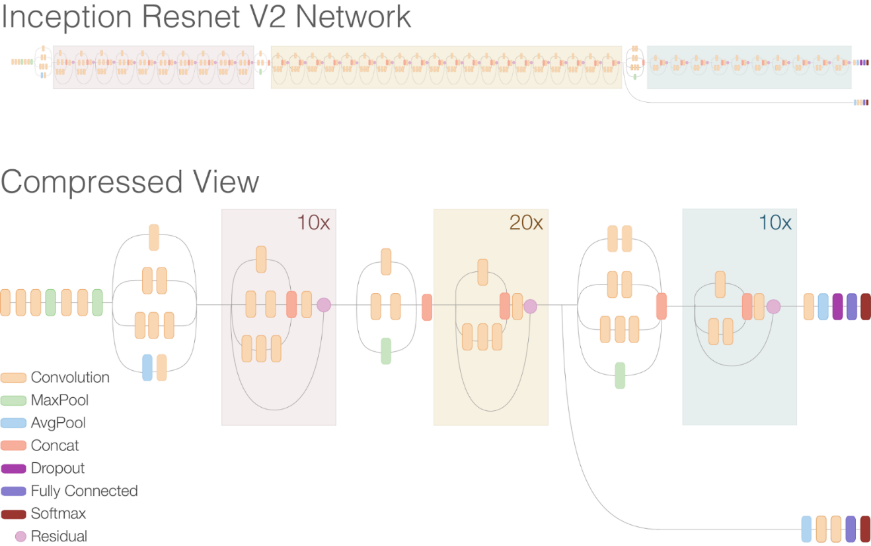
(układ warstw modelu)

Wśród różnych opracowań naukowych jako metody klasyfikacji najczęściej wykorzystuje się drzewo decyzyjne, maszynę wektorów nośnych (*SVM*), sztuczne sieci neuronowe czy dynamiczne sieci Bayesa [20]. W tej pracy zostanie użyta złożona sieć neuronowa bazująca na warstwach typu konwolucyjnego.

Duża część współczesnych architektur odznacza się nieliniową topologią. Sieci z rodziny „Inception” (opracowane przez zespół firmy „Google”) są oparte na modułach, których wejścia przetwarzane są przez kilka równoległych rozgałęzień konwolucyjnych [1]. W późniejszych krokach wejścia te są łączone. W skład sieci wchodzi stos modułów. Każdy z poszczególnych członów posiada trzy lub cztery rozgałęzienia. Cechą takiej budowy jest ułatwienie uczenia zależności kanałowych i przestrzennych. Każde z rozgałęzień różni się by móc nauczyć się innych cech.

Dodatkowym system wpływającym w znacznym stopniu na osiągnięcia okazała się technika pozwalająca dodawać połączenia szczątkowe do modelu. Sieci tej grupy są nazywane „ResNet” i zostały opracowane przez firmę „Microsoft”. Polega na ponownym wprowadzaniu do sieci reprezentacji z jej wcześniejszych etapów, co zapobiega traceniu informacji podczas kolejnych etapów przetwarzania danych [1].

Obie architektury przedstawione wyżej są bardzo wydajne, posiadają przy tym niskie koszty obliczeniowe. Podczas tworzenia tej pracy zostało przetestowanych wiele rozwiązań. Własny pomysł bazujący na ogólnie znanych schematach przynosił dobre wyniki, ale nie mogły się one równać się z uzyskiwanymi przez zespoły badawcze. Dlatego zdecydowano się użycie istniejącej architektury. Rozwiązaniem przynoszącym najlepsze wyniki okazał się model bazujący na budowie „Inception-ResNet v2” [31] który jest dostępny w bibliotece „Keras”. Schemat ideowy tego rozwiązania został pokazany na rysunku 5.1.



Rysunek 5.1. Schemat ideowy „Inception-ResNet v2” [32].

Cechuje się prostotą implementowania i bazuje na funkcji, która daje szereg dodatkowych opcji. Jest możliwość wczytania wag modelu pochodzących z procesu uczenia na zbiorze „ImageNet” [33], który posiada ogromne zasoby. Nie skorzystano z tej opcji, ponieważ nie otrzymano dzięki niej dobrych wyników. Nawet gdy znaczna część warstw została odblokowana w celu uczenia na zbiorze treningowym. Sieć neuronowa musiała być uczona od podstaw. Wybrane narzędzia posiadają funkcje pozwalające na odrzucenia ostatnich warstw w celu użycia innych dostosowanych poziomów końcowych. Pozwalają na implementację tej bazy do swojego problemu, tutaj klasyfikacji.

(dobór hiperparametrów modelu)

Na tym etapie ustawiono również wagi konkretnych klas by zrównoważyć ich różnorodność liczbową. Dobrane wartości odpowiadają logarytmowi naturalnemu ze stosunku liczbowego konkretnych klas. Takie podejście pozwala na uzyskanie dużo lepszej generalizacji. Jedną z innych zalet jest też ułatwiony proces śledzenia strat. Przydzielone wartości wpływają na straty informując podczas każdego przebiegu w dużo lepszy sposób o generalizacji sieci na danym etapie.

Większość wcześniej wspomnianych hiperparametrów była dobrana na podstawie wyników implementacji funkcji z biblioteki „Hyperopt” i biblioteki ułatwiającej optymalizację „Hyperas”. Obie z nich pozwalaj na pewną optymalizację doboru hiperparametrów. Ich głównym zadaniem jest przetworzenie modelu dla wcześniej dobranych opcjonalnych hiperparametrów i zwrócenie tych które po kilku przebiegach pozwoliły osiągnąć najlepszy wynik.

**Results**.

Analiza wyników to kłopotliwe zadanie, zwłaszcza gdy posiadane klasy mają różny poziom istotności. Są problemy, w których lepiej uzyskać gorszy wynik ogólny, ale najlepszy wynik z możliwych dla rozpoznania kilku szczególnych klas. W tabeli 6.2 zostały przedstawione dokładne wyniki klasyfikacji. Wstępne założenia wskazywały na osiągnięcie wyników skuteczności powyżej 90% dla każdej klasy co zostało spełnione.

Tabela 6.2. Wyniki klasyfikatora grup uderzeń serca.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grupa standardu *AAMI* | Skuteczność [%] | Precyzja [%] | Czułość [%] | Specyficzność [%] |
| N | 98,45 | 98,70 | 99,41 | 94,07 |
| S | 98,93 | 90,59 | 75,62 | 99,73 |
| V | 99,67 | 96,75 | 98,25 | 99,77 |
| F | 99,73 | 94,38 | 75,50 | 99,96 |
| Q | 99,95 | 99,38 | 100,00 | 99,95 |

Można stwierdzić, że ekstrakcja cech pozwoliła na otrzymanie wyników z małym poziomem błędu. Należy zauważyć, że niektóre klasy poradziły sobie gorzej. Grupa „F” osiągała najsłabsze wyniki ze wszystkich przed dodaniem zmodyfikowanych danych do zbiorów treningowych. Dokładny wynik klasyfikacji jest widoczny w macierzy pomyłek na rysunku 6.1. Do przeprowadzenia ewaluacji został wykorzystany tylko zbiór testowy. Przestrzeń i zakres kolorów wybrano w celu lepszego uwidocznienia zależności dla zbioru z dużymi dysproporcjami klasowymi.

Wyniki wskazują, że największą pomyłką jest klasyfikowane uderzeń nadkomorowych („S”) jako uderzeń normalnych („N”). Klasa „Q” uzyskała najlepsze wyniki. Tylko nieliczne próbki danych zostały rozpoznane jako inne klasy. Można wnioskować, że cechy charakteryzujące tą grupę są bardzo mocne i dobrze definiują przynależność. Macierz pomyłek pozwala również na wykazanie podobieństw. Zauważalna jest dużą liczbę pomyłek między klasą „S” i „N”. Taka zależność jest spotykana również w innych pracach. Musi to oznaczać podobieństwo obu grup.

W diagnostyce sygnałów *EKG* najważniejszym zadaniem jest rozdział grupy pobudzeń nadkomorowych i pobudzeń komorowych. W pracy udało uzyskać się liczbę tych błędów na bardzo niskim poziomie (6 pomyłek). Przekłada się to na około 0,29 % wszystkich próbek tych grup.

Klasę fuzji uderzeń („F”) można opisać jako zjawisko nakładania się na siebie typu komorowego oraz normalnego. Nie dziwi więc fakt istnienia błędów tej grupy dotyczących sklasyfikowania kilku próbek do klasy normalnej (N) i komorowej (V). Brak błędów w innych grupach dotyczących tej klasy możne oznaczać względnie dobre wyuczenie cech grupy przed badany model.

W celu dokonania lepszego porównania została stworzona tabela 6.1, która pokazuje procentowe wyniki wielu prac badawczych. Klasa uderzeń komorowych (V) wypadła bardzo dobrze na tle innych prac. Trochę słabiej wypadła grupa uderzeń nadkomorowych (S) ale nie odbiega znacząco od innych wyników. Główne miary statystyczne są tylko nieznacznie gorsze od dobrych wyników z ostatnich lat.

Tabela 6.1. Wyniki porównawcze klasyfikacji dla grupowania *AAMI* z wykorzystaniem bazy *MIT-BIH*.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Praca bad. | ACC | TPR | SPC | TPR (N) | (N) | ACC (S) | TPR (S) | (S) | ACC (V) | TPR (V) | (V) |
| [22] 2004 | 81,9 | - | - | 86,9 | 99,2 | 94,6 | 75,9 | 38,5 | 97,4 | 77,7 | 81,9 |
| [20] 2012 | 86,4 | - | - | 88,5 | 97,5 | - | 60,8 | 52,3 | - | 81,5 | 63,1 |
| [37] 2014 | 86,7 | - | - | 88,9 | 99,0 | - | 79,1 | 36,0 | - | 85,5 | 92,8 |
| [38] 2016 | 96,40 | 68,80 | 99,50 | - | - | - | - | - | - | - | - |
| [13] 2017 | 93,1 | - | - | 98,4 | 95,4 | - | 29,5 | 38,4 | - | 70,8 | 85,1 |
| [39] 2018 | 98,63 | 98,79 | 97,87 | - | - | - | - | - | - | - | - |
| [40] 2018 | - | - | - | - | - | 93,78 | 88,39 | 33,63 | 96,63 | 77,74 | 69,20 |
| [41] 2018 | 98,6 | 96,5 | 99,1 | 97,3 | - | 98,4 | 96,1 | - | 99,1 | 96,8 | - |
| [18] 2020 | 98,17 | 97,78 | 98,57 | 99,24 | - | 98,99 | 70,04 | - | 98,77 | 89,28 | - |
| Opracowanie własne | 98,37 | 98,54 | 95,08 | 99,41 | 98,70 | 98,93 | 75,62 | 90,59 | 99,67 | 98,25 | 96,75 |

Przytoczone prace posiadają różne procesy przetwarzania oraz klasyfikacji. Dotyczy to także liczby danych przekazanych na proces uczenia i zbiór testowy. Zmiana proporcji w znacznym stopniu wpływa na wyniki oraz dokładność przedstawionych zależności końcowych.

Choć analiza dotyczy ostatniego modelu osiągającego najlepsze wyniki należy wspomnieć o kilku zmianach, które w zauważalny sposób przyczyniły się do poprawy ewaluacji. Na początku zdjęcie po ekstrakcji cech zawierało tylko dane jednego sygnału. Dodanie drugiego zwiększyło skuteczność o średnio dwa punkty procentowe. Funkcja tworząca spektrogram posiada wiele parametrów. Początkowo w wyniku pomyłki zastosowano transformatę całego odcinka zamiast kroczącej transformaty Fouriera. Wyniki dla sieci głębokiej z małą ilością warstw poprawiły się. Sieci zaawansowane o dużej ilości warstw ukrytych przyniosły tylko nieznacznie lepsze wyniki.

Wartości miar statystycznych grupy „F” były bardzo słabe. Utworzeniu danych zmodyfikowanych, które w znaczny sposób zwiększyły liczność informacji w zbiorach treningowych zaobserwowano znaczą poprawę. Próba zwykłego powielenia danych tej grupy nie przyniosła pozytywnych rezultatów.

Początkowo ze względu na brak implementacji generatorów użyto mniejszej ilości danych. Wyniki nie były zadowalające, dużo lepiej model reagował na niezrównoważony zbiór treningowy, ale posiadający sporo większą ilość danych uczących. Prawdopodobnie charakter ekstrakcji cech i użycie głębokiej sieci neuronowej pozwoliło by na uzyskanie lepszych wyników, gdyby informacji przetwarzanych w procesie uczenia było znacznie więcej. Liczba danych dla grupy pobudzeń nadkomorowych („S”) jest mała co ma odwzorowanie w wynikach.

Model sieci oraz hiperparametry dobiera się do rozwiązywanego problemu. Zdarza się, że lepiej, gdy posiada on mniej wag do nauki, ponieważ sieć może zapamiętać określoną niską ilość zmiennych i dlatego skupi się na najważniejszych. Podczas analizy modelu można było zauważyć pozytywny wpływ na wynik podczas zwiększania liczby wag. Uzyskane wyniki są zadowalające przyrównując je do innych prac i wcześniejszych założeń początkowych.

**Conclusions and future work**

Przedstawiona implementacja algorytmu służy głównie do sprawdzenia możliwości i parametrów rozwiązania. Gotowy system, który może być używany w diagnostyce posiada trochę inne cechy i jest cały czas rozwijany bazując na nowych informacjach. Najlepsze rozwiązania posiadają dużą odporność na błędy i zakłócenia. Każdy pomiar sygnałów pochodzących od pacjenta jest uzależniony od jakości przeprowadzonego procesu. Wpływa na to badany oraz specjalista odpowiedzialny za przebieg badania. Pacjent może posiadać mimowolne skurcze mięśni, które występują w wielu chorobach lub też inne dolegliwości zmieniające parametry składu ciała. Dobrze jest, jeśli model odpowiedzialny za klasyfikację nauczył się tylko takich cech, które różnią typy uderzeń serca. Pozwala to na małą ingerencję znanych jak i nie znanych przyszłych zniekształceń. Używając takich narzędzi specjalista dokonujący diagnozy musi przeprowadzić własny osąd. Nie powinien mieć pełnego zaufania do wykorzystywanych narzędzi, zwłaszcza że prawdopodobnie nie zna szczegółów działania algorytmu. Bezgraniczne ufanie wynikom programu może być nie właściwe, jeśli ma się do czynienia ze zdrowiem i życiem innego człowieka. Badania sygnału EKG są od wielu lat ważnym krokiem diagnostycznym, dzięki temu istnieje wiele standardów używanych na skalę światową. Na każdym etapie starano się kierować wytycznymi albo dobrymi praktykami.

Wyniki można ocenić jako zadowalające. Zastosowanie zbioru walidacyjnego pozwoliło na lepszą separację danych testowych. Takie podejście powoduje, że część informacji podczas procesu doboru hiperparametrów przedostaje się tylko do zbioru, który jest wykorzystywany podczas procedury uczenia. Dlatego trenowanie na zbiorze treningowym, dobieranie hiperparametrów na zbiorze walidacyjnym, a przeprowadzanie testów na zbiorze testowym pozwala na większą generalizację i lepsze odwzorowanie warunków rzeczywistych. Jeszcze lepszym rozwiązaniem mogłoby być zastosowanie danych do fazy testowej, które pochodzą z zupełnie innej dystrybucji.

Dąży się do prostych, niewymagających dużej ilości zasobów rozwiązań. Przedstawiony klasyfikator do nich nie należy, choć nie można powiedzieć za dużo przed implementacją i praktycznym użytkowaniem. Przypuszczalnie czas potrzebny do przetworzenia może okazać się wystarczająco niski, a wyniki dostatecznie poprawne by rozwiązanie bazujące na rozbudowanym modelu sieci neuronowej zostało wdrożone.

Jednym z pomysłów na modyfikację pracy jest użycie dwóch zdjęć spektrogramów nie łącząc sygnałów w jeden obraz. W takim wypadku model posiadałby dwa wejścia, które łączyły by się po kilku blokach sieci neuronowej. Prawdopodobnie wyniki były by lepsze z tego względu, że w użytej bazie sygnały nie zawsze pochodzą z tych samych czujników.

Dużo lepsze wyniki można uzyskać zazwyczaj dzięki zastosowaniu zespołów modeli. Bloki powinny różnić się w znacznym stopniu, a jednocześnie uzyskiwać dobre wyniki. Różnorodność modeli pozwala na dostarczanie informacji nie dostępnych dla jednego typu sieci neuronowej. Połączenie takich modeli powinno nastąpić przy użyciu odpowiedniej średniej ważonej. Wagi ustalane są zależnie od osiąganych wyników typu sieci. Prace badawcze pokazują, że można uzyskać dobre efekty używając sieci rekurencyjnych. Ujemnym czynnikiem może być jedynie większe zapotrzebowanie na moc obliczeniową.

Ręczne dostrajanie hiperparametrów modelu nie ma sensu. Lepiej korzystać z automatyzowanych technik, pozwala to na szybsze i pewniejsze uzyskanie optymalizacji w tej dziedzinie.

Kilka występujących zakłóceń nie zostało usuniętych z przebiegów ze względu na ich powszechność we wszystkich sygnałach. Można sądzić, że ze względu na ich charakter nie mają zbytniego wpływu na wyuczone cechy, a tym samym wyniki przebiegów pochodzących z baz zewnętrznych, które prawdopodobnie nie posiadają tych zakłóceń.

Na samym początku z powodu błędu, spektrogram był tworzony na całym przebiegu. Transformata Fouriera obejmowała cały odcinek każdego uderzenia serca. Uzyskane wyniki były bardzo dobre. Można sądzić, że cechy, które decydują o przynależności do konkretnej grupy zostały odnalezione przy użyciu zwykłej transformaty Fouriera.

Gdy dane zostaną podzielone losowo na zbiory istnieje ryzyko, że sieć dopasuje się do cech zapisu pacjenta. Jest to problem, który ciężko zweryfikować. Należy brać pod uwagę kroki przetwarzania i ekstrakcji cech, które zminimalizowały możliwość wystąpienia tego problemu przez zanik pewnych informacji sygnału. Utworzony projekt jest skonstruowany w taki sposób by przy użyciu prostych czynności dojść do próbki, która jest błędnie sklasyfikowana i dowiedzieć się na czym ten błąd polegał.

Podczas pracy nad siecią neuronową był pomysł na użycie części wag gotowego modelu, który był uczony na znanej bazie obrazów. Pozwalając jedynie na zmianę kilku warstw końcowych by zostawić wyuczone przekształcenia i skrócić w znaczny sposób przebieg uczenia. Niestety wyniki były dużo gorsze niż podczas procesu uczenia całego modelu, począwszy od randomizowanych wag. Ciekawym rozwiązaniem jest wykorzystanie dodatkowo informacji o interwałach RR, byłyby przetwarzane podczas tego samego procesu uczenia. Zrealizować to można tworząc drugie wejście na dane, które przyjmowało by tensor z tymi informacjami.

Zastosowane podejście w oryginalny sposób podchodzi do ekstrakcji cech. Tworzenie jednego obrazu z dwóch sygnałów zwiększa nakłady obliczeniowe tylko w minimalnym stopniu. Nie zwiększa to liczby pikseli obrazu, czyli czas na wytrenowanie sieci nie powinien wzrosnąć. Uzyskanie dobrych wyników dla grupy „V” i „S” jest bardzo ważne. Jednym ze sposobów jest ustawienie większej wagi podczas procesu uczenia by błędy dotyczące tych klas dawały większe wartości straty. Algorytm skupiałby się bardziej na tych problemach by zminimalizować ogólną wartość. Takie podejście zostało zaimplementowane i przyniosło rezultaty pozytywne.

Zastosowane etapy przetwarzania pozwalają na szybkie przetworzenie i włączenie do zbioru treningowego danych pochodzących z innych źródeł. Zapisy dotyczące pacjentów mogą posiadać również cechy indywidualne, które objawiają się w sygnale. Mogą istnieć przypadki ewolucji serca sklasyfikowane nie na podstawie cech grupy, ale na podstawie zależności pochodzących od badanej jednostki. Tak poważny błąd dyskwalifikuje rozwiązanie z rzeczywistego użycia. Jeśli zostałby wykryty należałoby zastosować dodatkowe źródła danych i jeśli to możliwe zmienić procesy ekstrakcji cech by skupić się na wyszczególnieniu odpowiednich zależności.

System, który działałby w rzeczywistości może zbierać nowe informacje w celu poprawy skuteczności. Wykorzystuje się do tego trenowanie ciągłe. Jednym z problemów systemu uczącego się na nowych danych jest różny sposób zapisu sygnałów. Każdy wymaga innego podejścia. Są też inne zagadnienia związane w mniejszym lub większym stopniu z tematem tej pracy, które można diagnozować na podstawie sygnału EKG. Przypuszczalnie każde nowe dane mogły by wnieść dodatkowe informacje pomagające usprawnić ten klasyfikator.