**AKADEMIA GÓRNICZO-HUTNICZA**

**im. Stanisława Staszica w Krakowie**

**Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej**

**Praca dyplomowa  
inżynierska**

**Michał Rola**

*Imię i nazwisko*

**Automatyka i Robotyka**

*Kierunek studiów*

**Automatyczne identyfikowanie gatunków ptaków  
na podstawie nagrań ich wokalizacji**

*Temat pracy dyplomowej*

**Dr Inż. Andrzej Izworski**

*Promotor pracy*

……………………

*Ocena*

**Kraków, rok 2023/2024**

Spis treści

[1 Wstęp 3](#_Toc152147269)

[2 Przedstawienie problemu badawczego 3](#_Toc152147270)

[2.1 Mel Spektrogramy 3](#_Toc152147271)

[3 Zbiór danych próbek audio 4](#_Toc152147272)

[3.1 Akwizycja próbek audio 4](#_Toc152147273)

[3.2 Przetwarzanie cyfrowe próbek audio 5](#_Toc152147274)

[3.3 Przygotowanie zbiorów danych 7](#_Toc152147275)

[4 Wybór i konstrukcja sieci neuronowej 8](#_Toc152147276)

[4.1 Przegląd dostępnych rozwiązań 8](#_Toc152147277)

[4.2 Wykorzystane technologie oraz narzędzia 8](#_Toc152147278)

[4.3 Wybór modelu oraz uzasadnienie 8](#_Toc152147279)

[4.4 Projektowanie oraz uczenie modelu 8](#_Toc152147280)

[5 Prezentacja i analiza wyników 8](#_Toc152147281)

[6 Podsumowanie 8](#_Toc152147282)

[Bibliografia 9](#_Toc152147283)

[Dodatek 1 – Dziedzinowy słownik pojęć 9](#_Toc152147284)

[Dodatek 2 – Przykładowe próbki wokalizacji 9](#_Toc152147285)

# Wstęp

(1.5 strony, czego dotyczy praca)

# Przedstawienie problemu badawczego

(3-4 strony)

## Mel Spektrogramy

Mel Spektrogramy to Spektrogramy posiadające Skale Mela[1] na osi OY zamiast liniowo rosnącej częstotliwości. Skala Mela powstała dzięki badaniom polegającym na ocenie przez słuchaczy do jakiej częstotliwości należy dany ton. Jest więc to miara subiektywna i posiada wiele równań zależnie od interpretacji tych danych, jednak znacznie lepiej przedstawia w jaki sposób ludzkie ucho postrzega dźwięki. W pracy tej wykorzystane zostały wzory Slaney’a (2.1) oraz HTK (2.2).

(2.1)

(2.2)

# Zbiór danych próbek audio

Stworzenie zbioru treningowego zawierającego dużo wysokiej jakości próbek jest jednym z ważniejszych i bardziej czasochłonną częścią tworzenia dobrego modelu.

## Akwizycja próbek audio

Próbki audio pobrano ze strony https://xeno-canto.org. Jest to strona, której użytkownicy z całego świata wspólnie zbierają, identyfikują oraz dzielą się doświadczeniem dotyczącym nagrań zwierząt. Poza nagraniami wokalizacji ptaków, które stanowią lwią część zbiorów strony, można znaleźć również nagrania odgłosów koników polnych oraz nietoperzy.

Postanowiono stworzyć zbiór danych składającą się z wokalizacji 10 gatunków ptaków, po 50 nagrań na gatunek dla zbioru treningowego oraz po 10 nagrań na gatunek dla zbioru testowego.

Przy doborze nagrań do zbioru treningowego skupiono się na tym, aby posiadały jak najmniej zakłóceń i żeby zawierały tylko odgłosy jednego gatunku. Ponadto nagranie musiało być nagrane z częstotliwością próbkowania równą co najmniej 32.0kHz oraz przepływnością stałą na poziomie minimum 320kb/s (przykładowy fragment kształtu przebiegu sygnału przedstawia rysunek 3.1). Nagrania do zbioru testowego przeszły mniej rygorystyczną selekcję w celu sprawdzenia, jak model poradzi sobie z próbkami o jakości bliższej realnym warunkom.

A screenshot of a sound wave

Description automatically generated

Rysunek 3.1 Przykładowy fragment kształtu przebiegu sygnału przed obróbką [źródło: opracowanie własne]

Po zebraniu wystarczającej ilości nieprzetworzonych nagrań wyizolowano wokalizacje poszczególnych gatunków. Uczono model na nagraniach posiadających po 1 wokalizacji o różnej długości (od 0,2-1,0 sekundy), z czego większość próbek trwa 0,3 sekundy. Zrezygnowano z jakości stereo. Dalsze prace prowadzono na sygnale mono w celu zmniejszenia objętości plików, przyśpieszenia późniejszej obróbki oraz aby model nie musiał analizować dwóch Mel Spektrogramów zawierających bardzo zbliżone do siebie sygnały (przykładowy kształt wyizolowanego sygnału zamieszczono na rysunku 3.2).

A screen shot of a sound wave

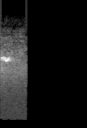
Description automatically generated

Rysunek 3.2 Kształt sygnału z rysunku 3.1 po wyizolowaniu wokalizacji ptaka oraz zmniejszeniu ilości kanałów do jednego [źródło: opracowanie własne]

Po zebraniu odpowiedniej ilości wstępnie przygotowanych próbek można było przejść do ich dalszego przetwarzania.

## Przetwarzanie cyfrowe próbek audio

Z pomocą biblioteki Librosa przetworzono próbki audio z formatu WAV  
(ang. *Waveform Audio Format*) na obrazy Mel Spektrogramów. Tak przetworzone obrazy zostały zapisane w formacie PNG (ang. *Portable Network Graphics*), który posiada bezstratną kompresję. Obrazy zapisano w odcieniach szarości, w celu ograniczenia zajmowanego przez nie miejsca, ponieważ zamiast trzech kanałów na kolory wymagany jest tylko jeden.  
Przykład formatu obrazu zastosowanego do uczenia sieci neuronowej przedstawiono na rysunku 3.3 oraz 3.4.



Rysunek 3.3 Zarys Mel Spektrogramu sygnału z rysunku 3.2 z wykorzystaniem równań Slaney’a  
[źródło: opracowanie własne]

A black and white photo of a light

Description automatically generated

Rysunek 3.4 Zarys Mel Spektrogramu sygnału z rysunku 3.2 z wykorzystaniem równania HTK  
[źródło: opracowanie własne]

## Przygotowanie zbiorów danych

Kod odpowiedzialny za tworzenie zbioru danych przedstawiono poniżej.



Tak przygotowane dane były gotowe do wprowadzenia do modelu.

# Wybór i konstrukcja sieci neuronowej

## Przegląd dostępnych rozwiązań

## Wykorzystane technologie oraz narzędzia

## Wybór modelu oraz uzasadnienie

## Projektowanie oraz uczenie modelu

# Prezentacja i analiza wyników

# Podsumowanie

(1.5 co zostało zrobione, co nie wyszło i dlaczego, co można poprawić)

# Bibliografia

[1] <https://librosa.org/doc/main/generated/librosa.mel_frequencies.html>

[2]

Tyle pozycji, aby przejść na kolejną stronę.

# Dodatek 1 – Dziedzinowy słownik pojęć

# Dodatek 2 – Przykładowe próbki wokalizacji

(link)