



**Politechnika  
Śląska**

**Wydział Automatyki, Elektroniki  
i Informatyki**

Systemy Interaktywne i Multimedialne  
Projekt  
Detekcja emocji w głosie

Natalia Stręć Jakub Kula, Paweł Wójtowicz,

Gliwice 2023

# Spis treści

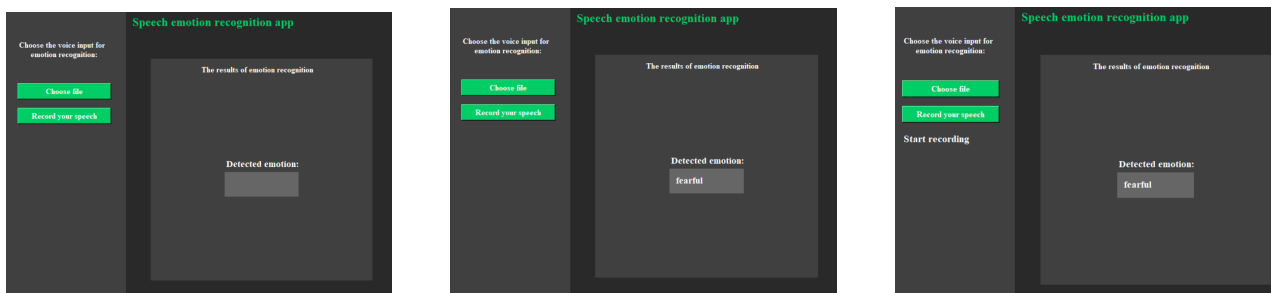
<b>1</b>	<b>Analiza wyników i wnioski</b>	<b>2</b>
1.1	UI . . . . .	2
1.2	Dokładność sieci . . . . .	2

# 1 Analiza wyników i wnioski

Głównym celem na maj było stworzono UI pozwalające na wybranie pliku dźwiękowego oraz nagranie wiadomości głosowej, używając Tkinter, z którego zostają wyciągane parametry służące do predykcji emocji. Celem dodatkowym była ciągła praca nad siecią nerwonową w celu poprawy dokładności na zbiorze testowym.

## 1.1 UI

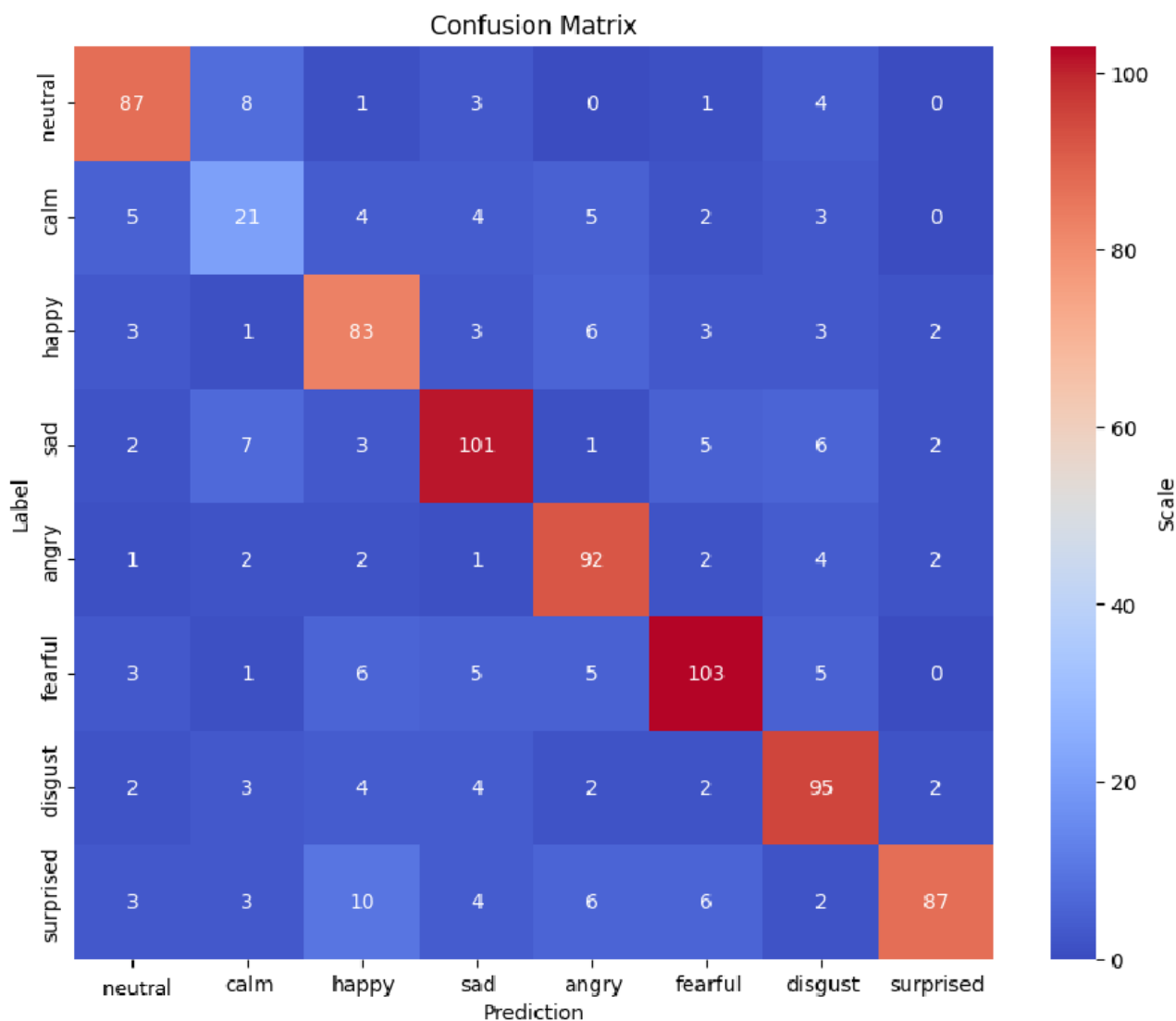
UI jest zaprojektowane w sposób intuicyjny, umożliwiając użytkownikom łatwe dodanie pliku dźwiękowego oraz nagranie własnej wiadomości przy użyciu wbudowanego mikrofonu.



Prace nad interfejsem użytkownika są w toku. Dotychczas stworzono pierwszy zarys UI, który umożliwia użytkownikom dodawanie plików dźwiękowych oraz nagrywanie wiadomości za pomocą wbudowanego mikrofonu. Interfejs jest zaprojektowany tak, aby był intuicyjny i łatwy w obsłudze. Na obecnym etapie prace nad UI są zaawansowane w 80%. Planowane są dalsze udoskonalenia, w tym integracja z modelem analizy emocji oraz poprawa responsywności i estetyki interfejsu. Celem jest zapewnienie użytkownikom wygodnego i efektywnego narzędzia do interakcji z systemem.

## 1.2 Dokładność sieci

W ramach prac nad dokładnością modelu, zostały wykonane testy optymalizacji hiperparametrów sieci. Proces dostosowywania parametrów modelu jest kluczowym krokiem w optymalizacji jego wydajności. Obejmuje to fine-tuning hiperparametrów, takich jak współczynnik uczenia, liczba neuronów w warstwach ukrytych oraz liczba epok treningowych. Dotychczasowe testy sugerują, że dostosowanie tych parametrów może znacząco poprawić dokładność modelu w rozpoznawaniu emocji. Wykazały one, że zwiększeni skomplikowania modelu, spowodowało zmniejszenie biasu, co poskutkowało znacznie lepszymi wynikami krosvalidacji. Obecny najlepszy model osiąga wartość 76% dokładności na zbiorze testowym.



Można zauważyć, że nowy model znacząco lepiej radzi sobie z klasyfikacją emocji w głosie. Różnica w ilości cech w zbiorze testowym dla emocji "calm" wynika z wybranych zbiorów nagrań głosowych. Pierwszy z nich posiadał 8, a drugi, większy, tylko 7.

Projekt jest realizowany zgodnie z harmonogramem oraz nie wprowadzono w tym etapie żadnych znaczących zmian odbiegających od niego.