



**Politechnika
Śląska**

**Wydział Automatyki, Elektroniki
i Informatyki**

Systemy Interaktywne i Multimedialne
Projekt
Detekcja emocji w głosie

Natalia Stręk Jakub Kula, Paweł Wójtowicz,

Gliwice 2023

Spis treści

1	Cele i zadania z okresu maj	2
2	Opis zadań przyjętych do realizacji	2
	Stworzenie UI	2
	Poprawa dokładności sieci	3
3	Zgodność z harmonogramem	3
4	Czy wprowadzono znaczące zmiany do projektu?	4

1 Cele i zadania z okresu maj

Głównym celem na maj było stworzenie UI pozwalającego 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 nerwową w celu poprawy dokładności na zbiorze testowym.

2 Opis zadań przyjętych do realizacji

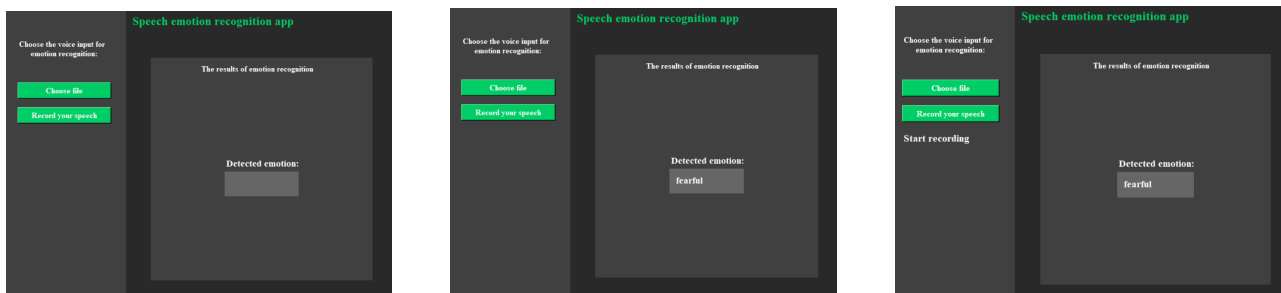
Do tego momentu zostały zrealizowane następujące zadania

- Zgromadzenie i przetworzenie danych - dane w postaci krótkich nagrań dźwiękowych reprezentujące różne emocje,
- Wybór oraz ekstrakcja cech,
- Wybór architektury sieci, hiperparametrów, podział danych na zbiór uczący oraz trenin-gowy,
- Trenowanie modelu i optymalizacja parametrów modelu,
- Ocena i walidacja modelu.

A następujące zadania zostały przyjęte do realizacji:

Stworzenie UI

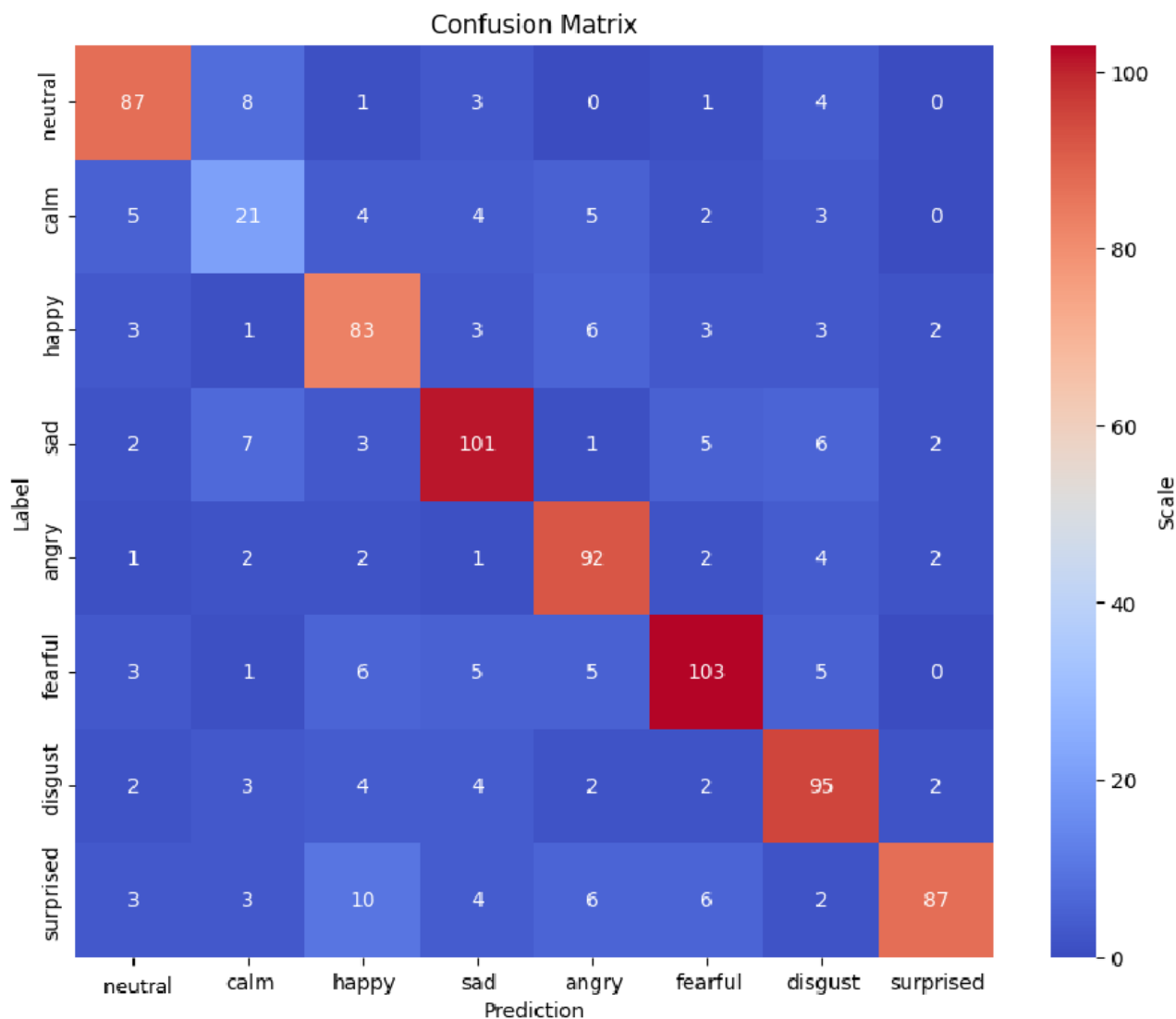
UI jest zaprojektowane w sposób intuicyjny, umożliwiając użytkownikom łatwe dodanie pliku dźwiękowego oraz nagranie własnej wiadomości przy użyciu wbudowanego mikrofonu. Aplikacja przyjmuje do dalszej analizy jedynie pliki z rozszerzeniem .wav oraz .mp3. Poniżej zamieszczono efekty dotychczasowej pracy. Prace nad interfejsem użytkownika są w toku. Do-



tychczas stworzono pierwszy zarys UI. Do nagrania pliku dźwiękowego wykorzystano bibliotekę sounddevice, w której można było ustawić m.in. odpowiednią częstotliwość próbkowania. 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.

Poprawa dokładności 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. Dostosowywanie parametrów w celu poprawy wydajności modelu jest zadaniem najbardziej czasochłonnym, które wymagało około tygodnia pracy. Projekt w ok. 90% jest gotowy.

3 Zgodność z harmonogramem

Projekt jest realizowany zgodnie z harmonogramem.

4 Czy wprowadzono znaczące zmiany do projektu?

Nie wprowadzono w tym etapie żadnych znaczących zmian odbiegających od harmonogramu projektu.