Processing and Recognition of Audio Signals

Mariia Slobodian

Wallee AG, Winterthur, Switzerland

Mykola Kozlenko

SoftServe, Austin, USA

Introduction

- Compared to all other animals, dogs have been in close relationships with humans for a very long period of time.
- A wide range of jobs in the field of modern human-machine interfaces has used automatic speech recognition systems during the past ten years.
- The increase in the quantity and availability of high-performance computing has played a crucial role in recognizing the emotional state of dogs by the acoustic signals.

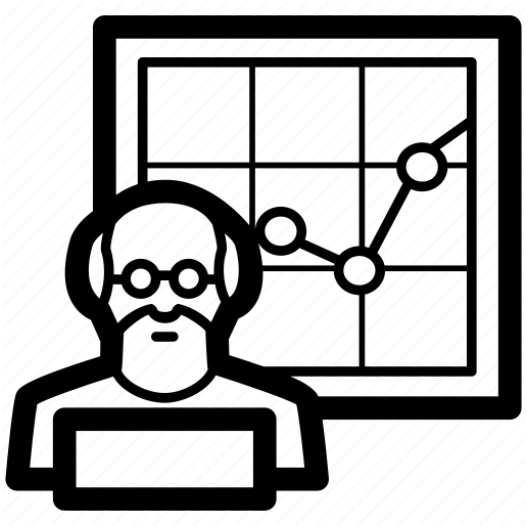
Problematic formulation

- Despite the significance of the issue, it should be noted that currently there is no comprehensive theory that investigates how the dogs' emotions link to the characteristics of the acoustic signals.

- The main challenge is that aggression and playfulness are generally active emotions and it is difficult to get fairly neutral vocalizations.

- Eventually, the best accuracy achieved so far is insufficient for use in real-world applications.

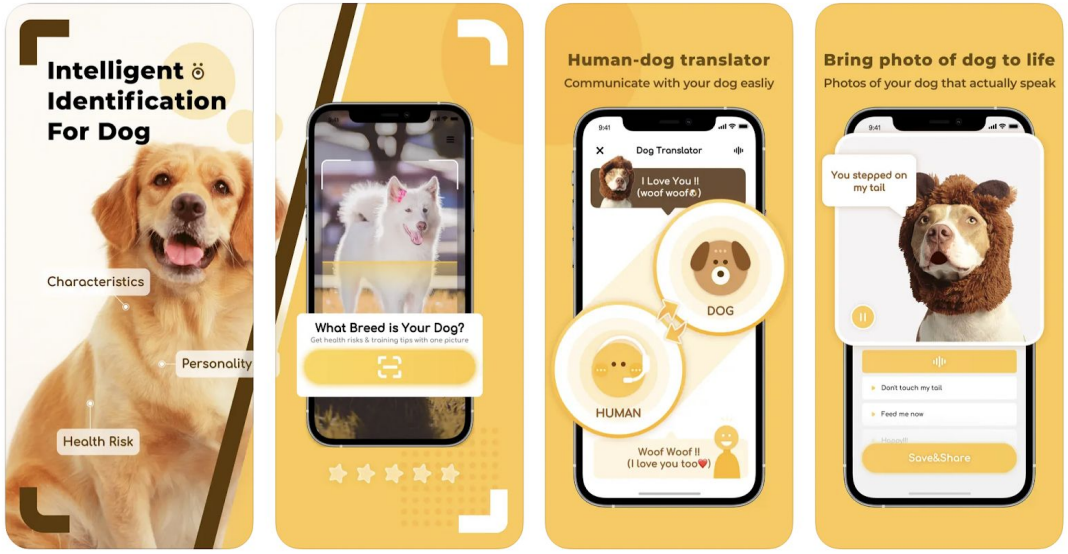
Similar scientific researches



Human speech emotion recognizer - **84.83%**

Human voice recognizer (male, female) - **97.6%**

Similar existing products



DogSnap

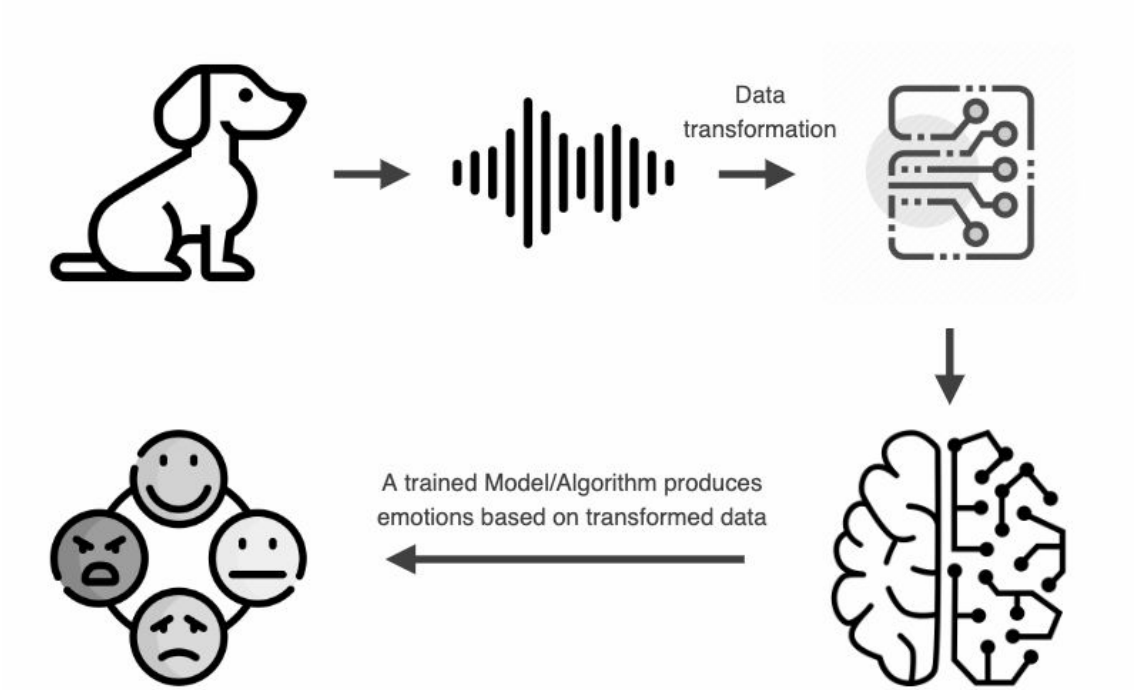
Petpuls

Happy Pets

TailTalk

Audio Signal Processing Systems

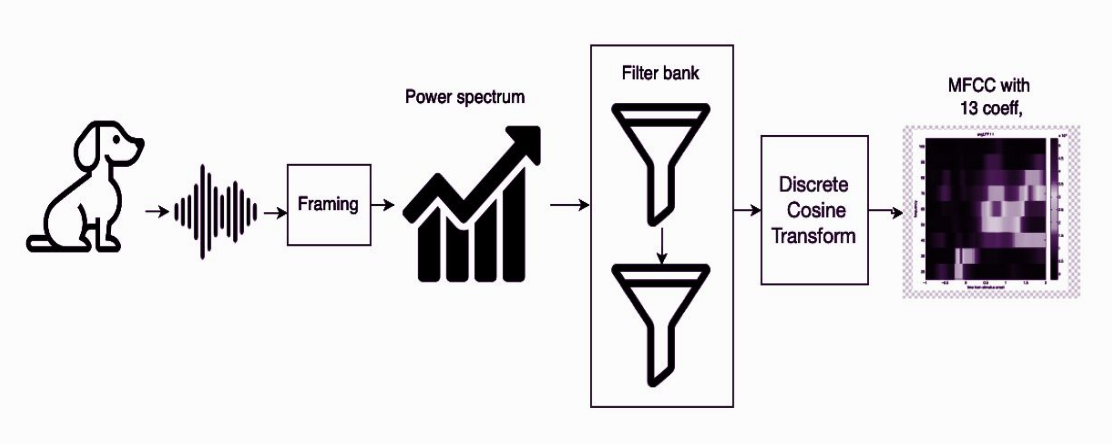
We studied the performance of two main methods with common data processing strategies: EDS and MFCC.

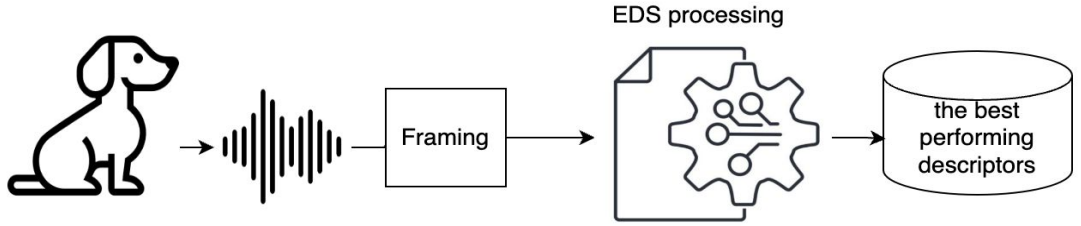


MFCC vs EDS

At the end of this process, the **MFCC** provides us with 13 useful coefficients that we use with proper machine learning algorithms

EDS iteratively generates descriptors for generation and could mark the best-performing descriptors from one generation as seeds for subsequent ones





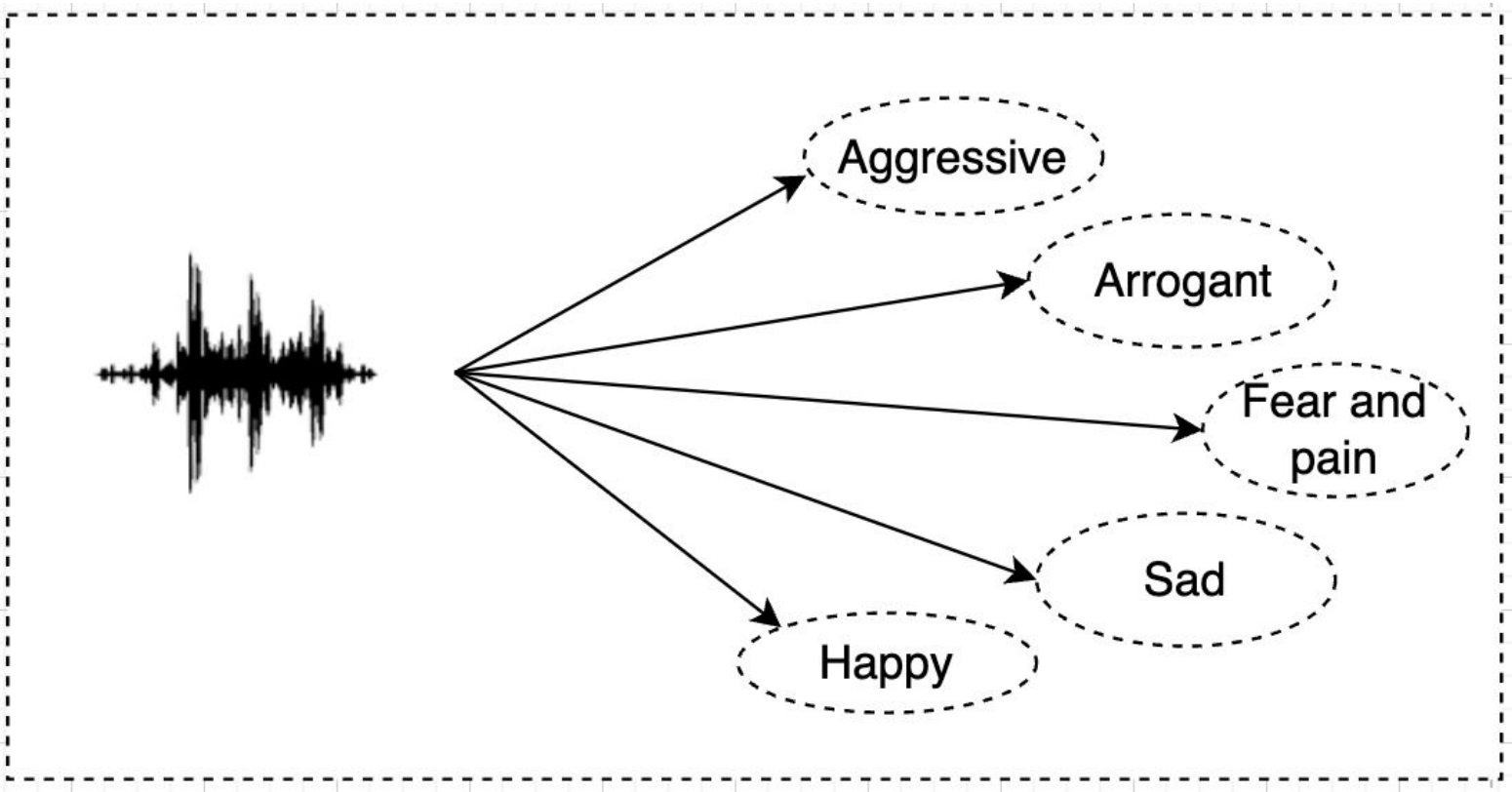
Audio features

The analysis of the dynamic movement of the vocal folds was physically challenging. To ease it, we identified the vocal characteristics by the following indicators of the voice:

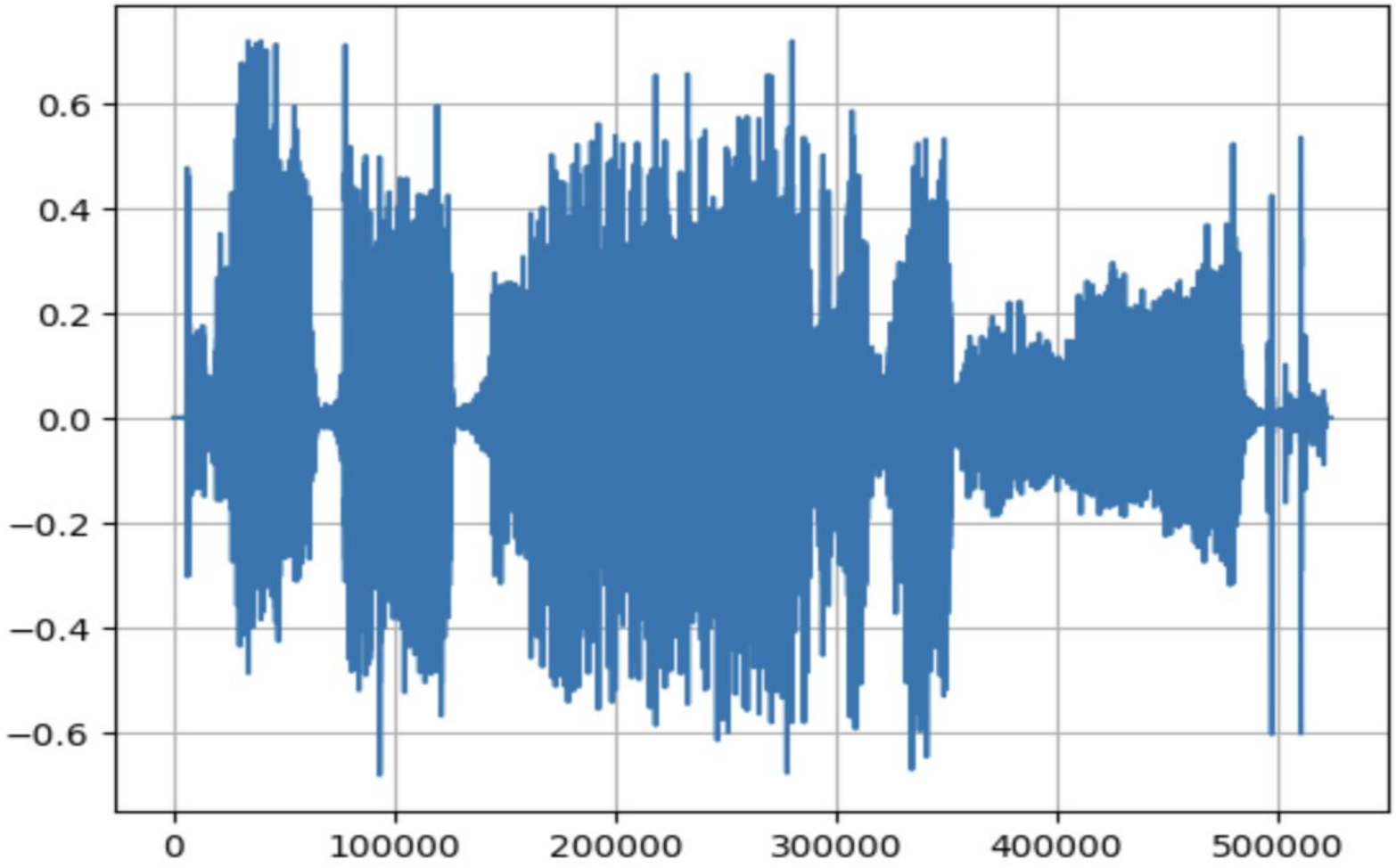
- Spectral composition
- Amplitude
- Intensity
- Timbre
- Loudness

Emotion classes

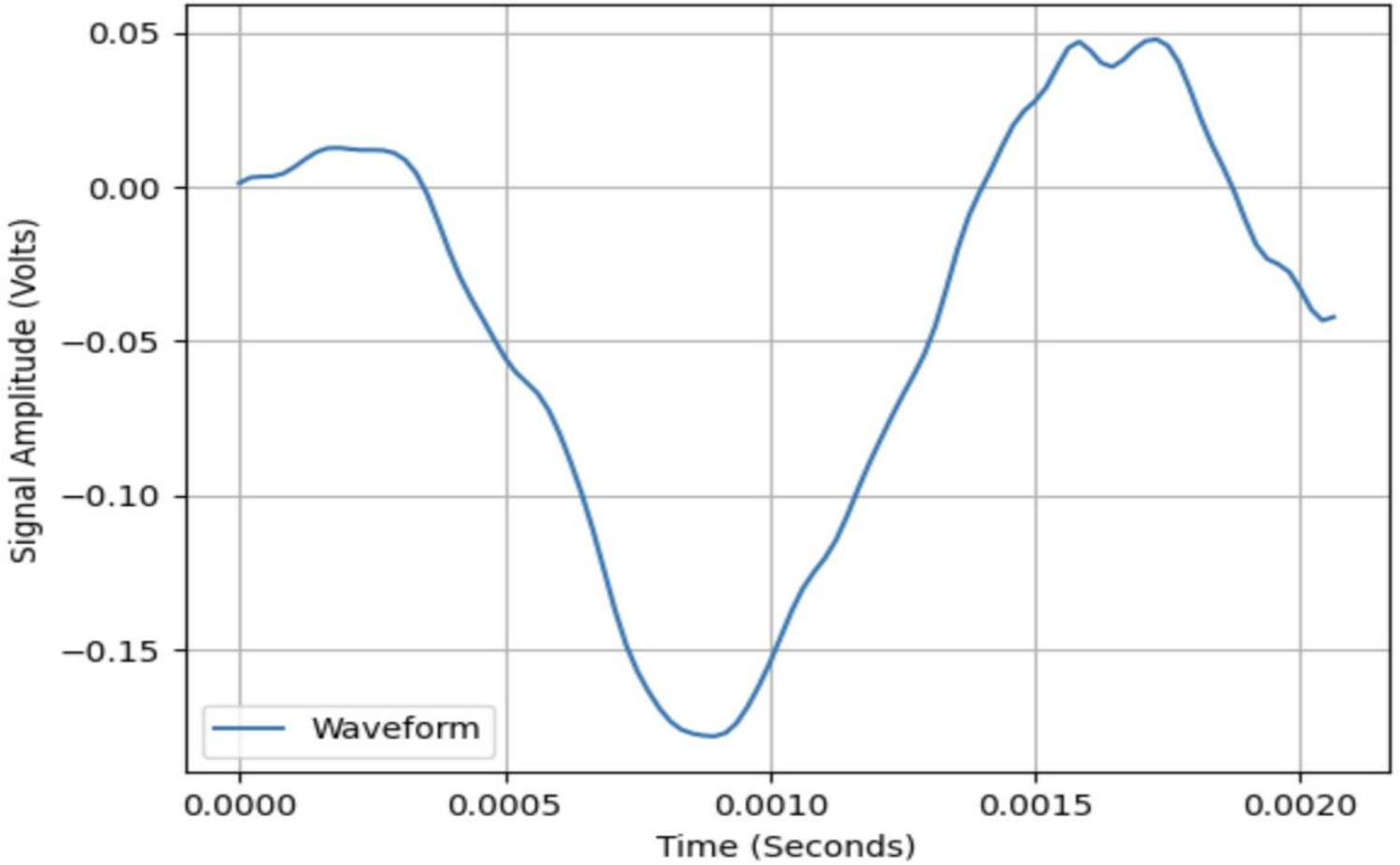
We determined that the ideal emotion sample has five types of feelings which are fewer compared to the human measurements.



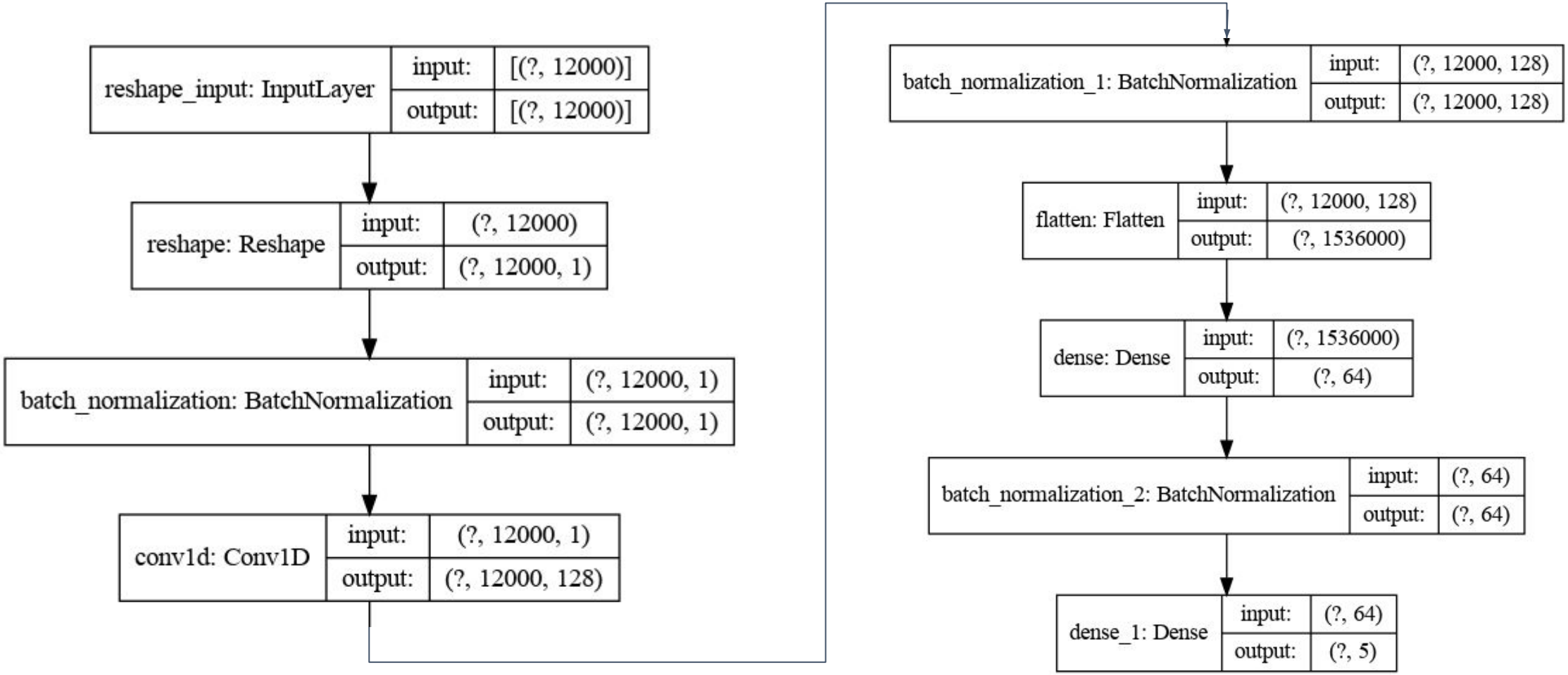
Audio representation



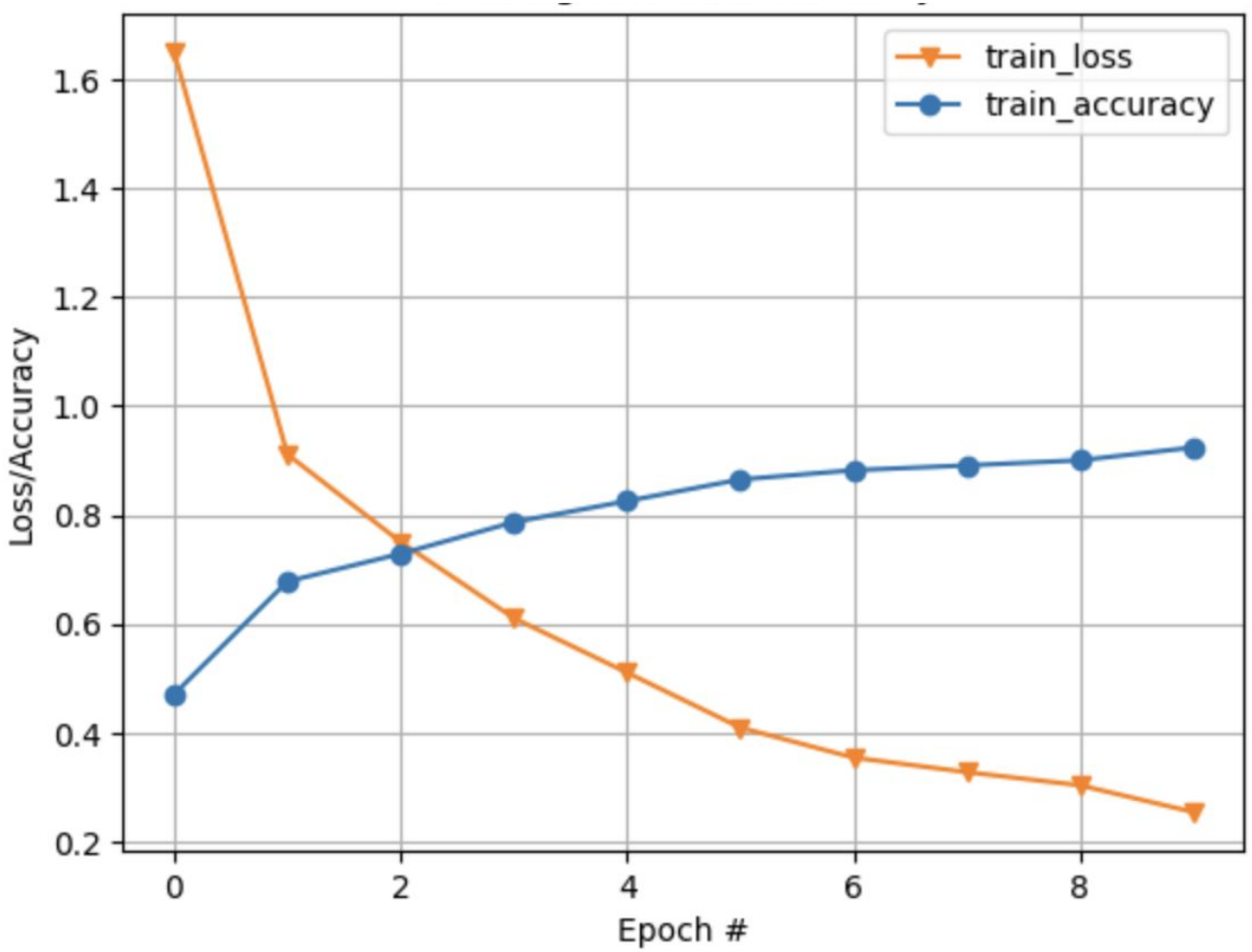
Signal waveform in time view



Model algorithm



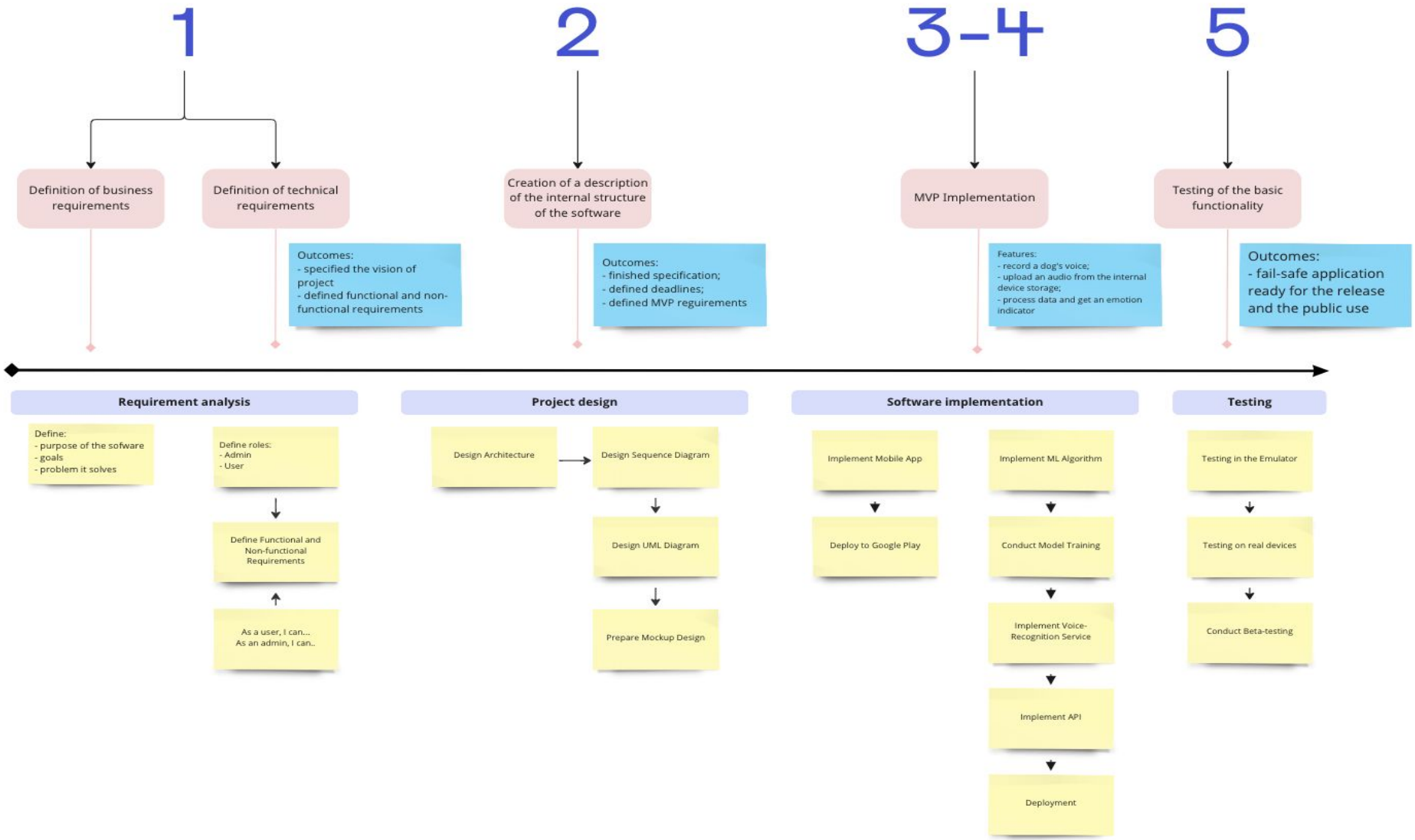
Training loss and accuracy



Model results

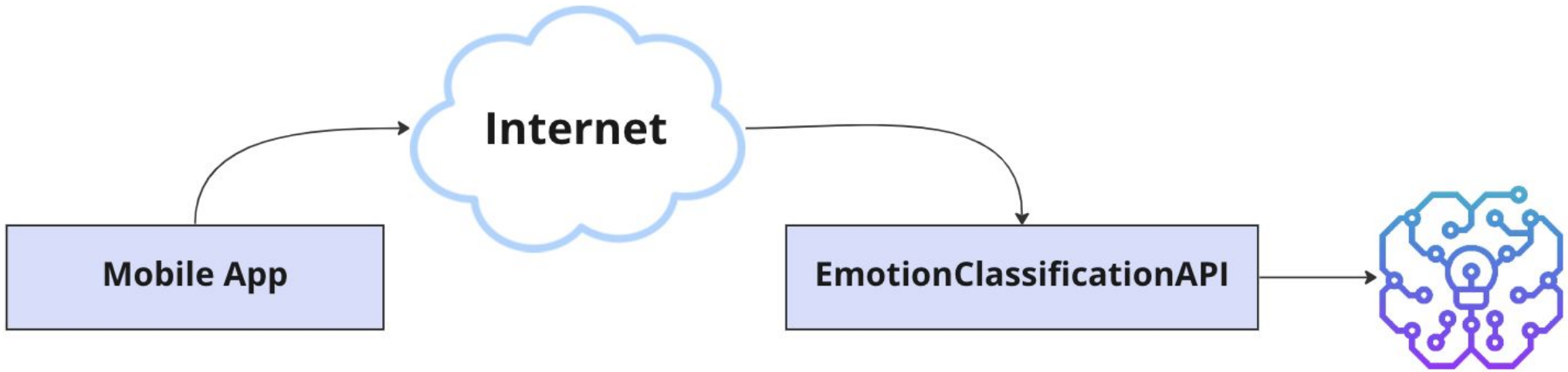
	precision	recall	f1-score	support
aggressive	0.73	0.81	0.77	280
arrogant	0.70	0.61	0.65	288
fear_and_pain	0.66	0.93	0.77	302
happy	0.80	0.54	0.64	329
sad	0.76	0.74	0.75	301
accuracy			0.72	1500
macro avg	0.73	0.73	0.72	1500
weighted avg	0.73	0.72	0.72	1500

Implementation roadmap

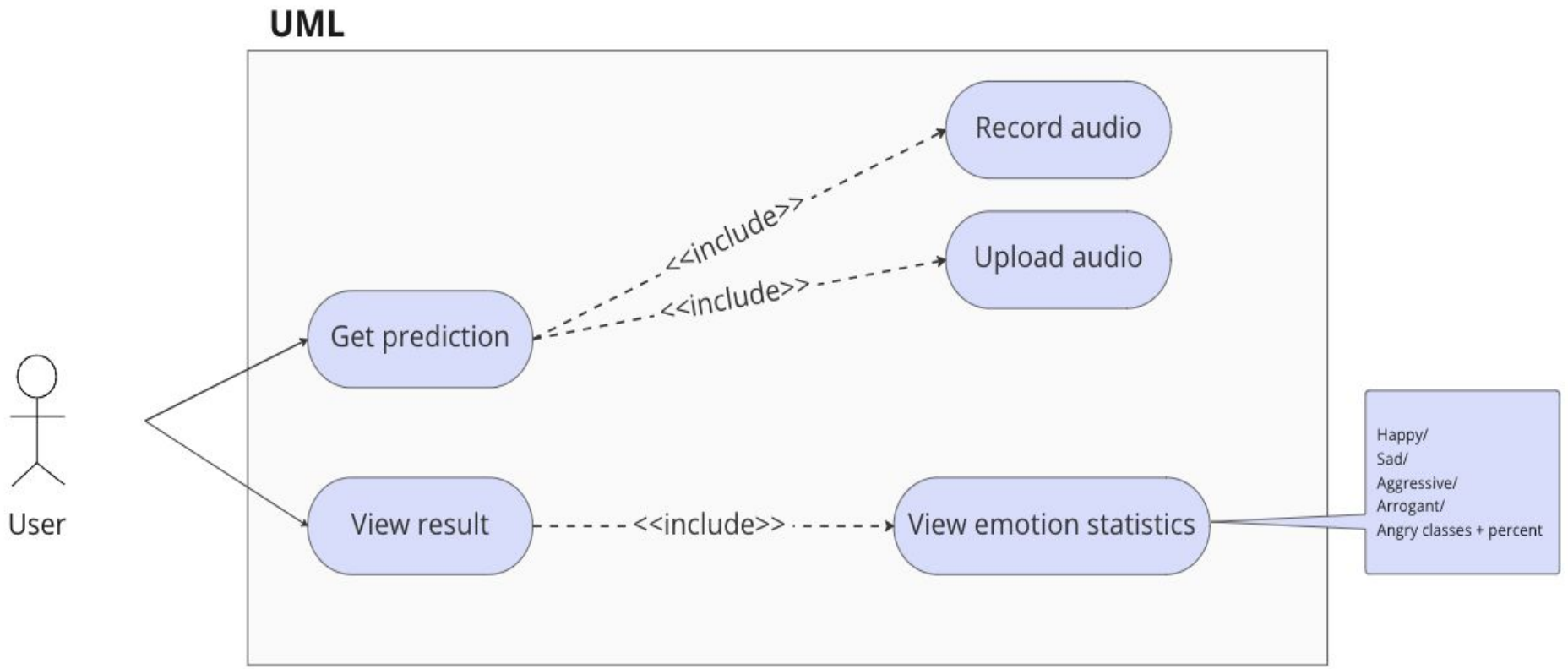


Technology

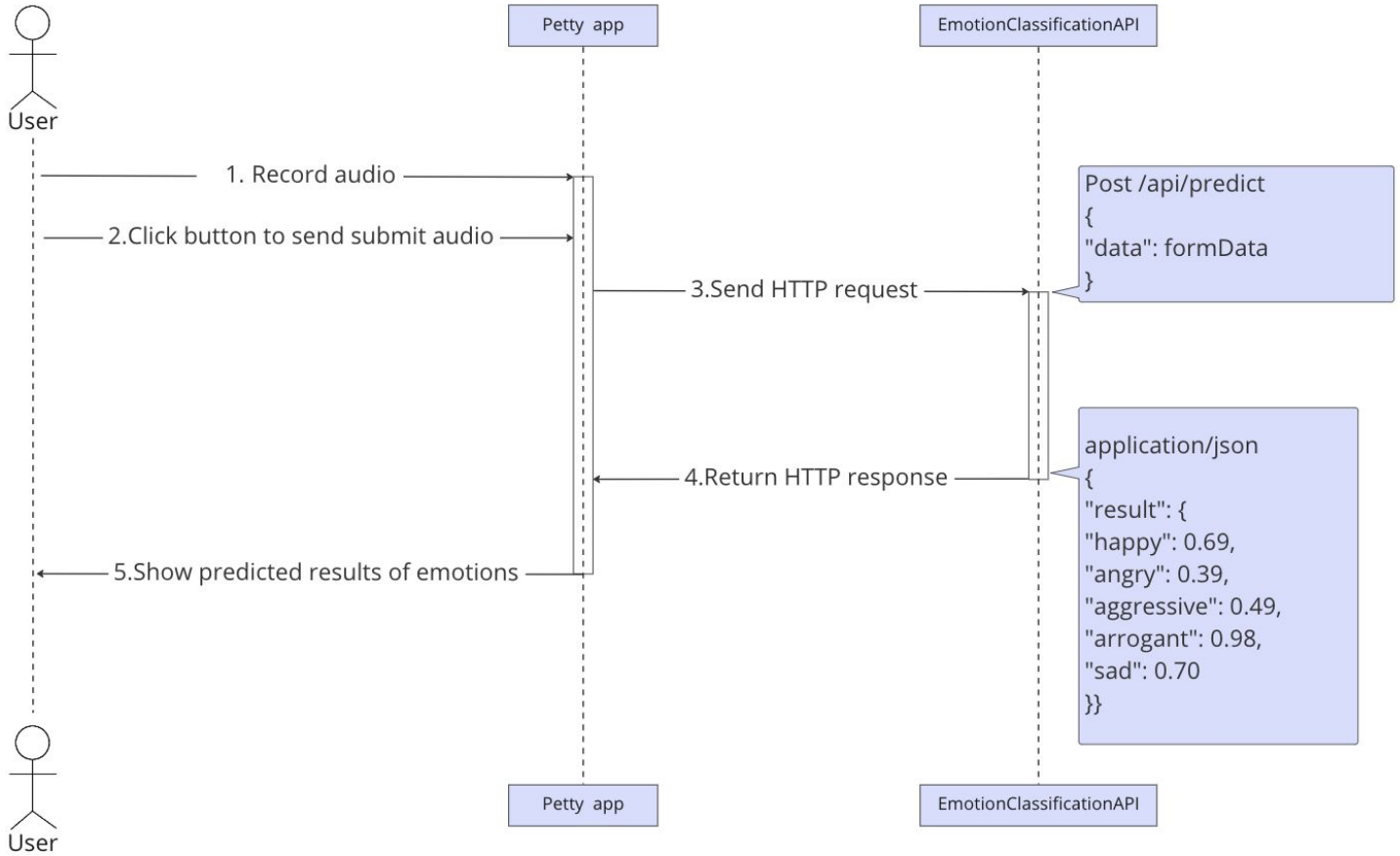
Architecture



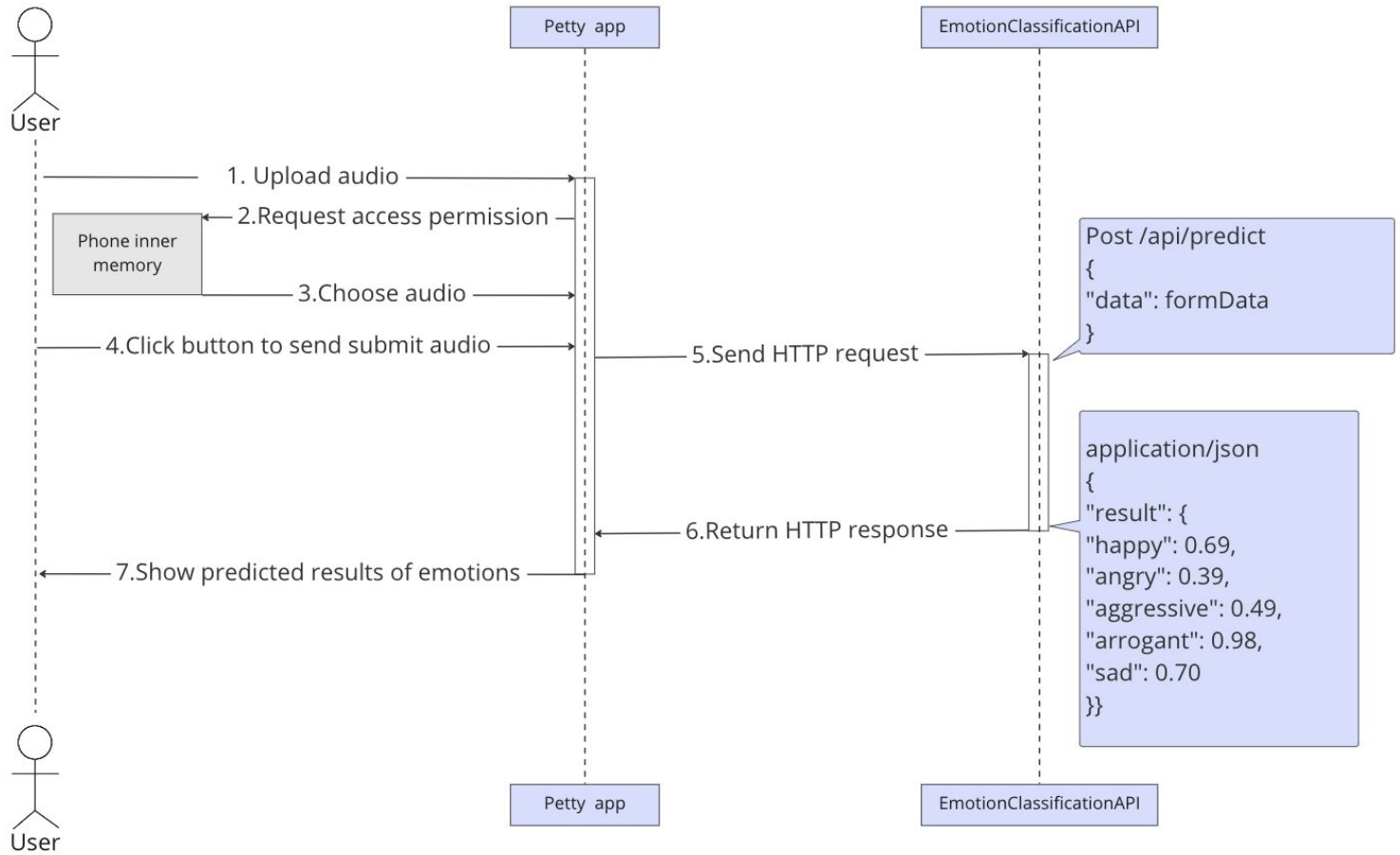
UML diagram


UML
Record audio
Upload audio
<<include>>
<<include>>
Get prediction
User
View result
<<include>>
View emotion statistics
Happy/
Sad/
Aggressive/
Arrogant/
Angry classes + percent

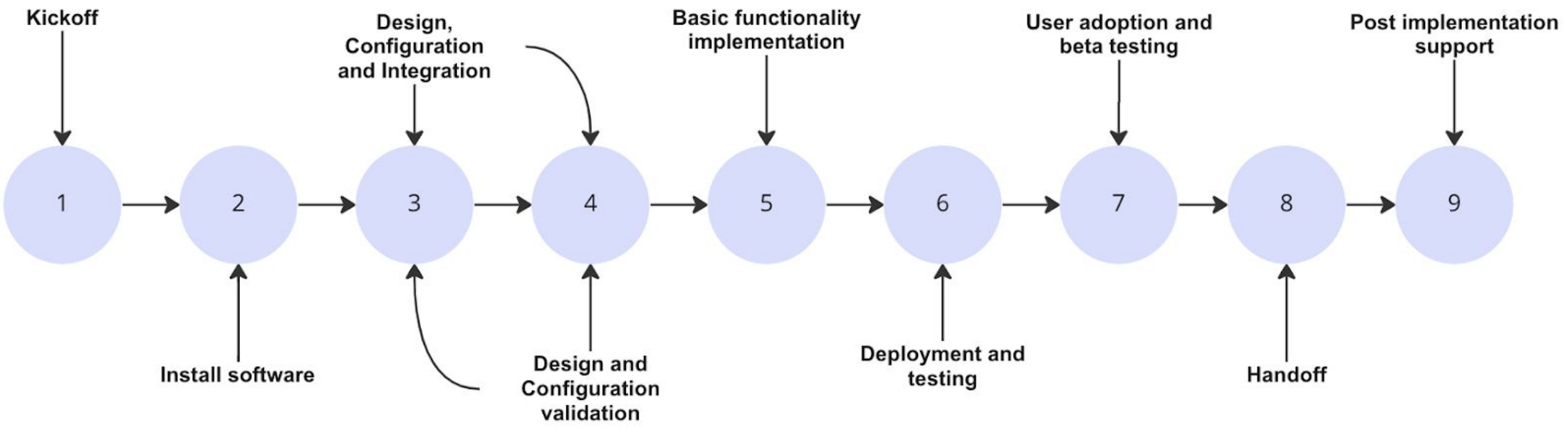
Sequence diagram I



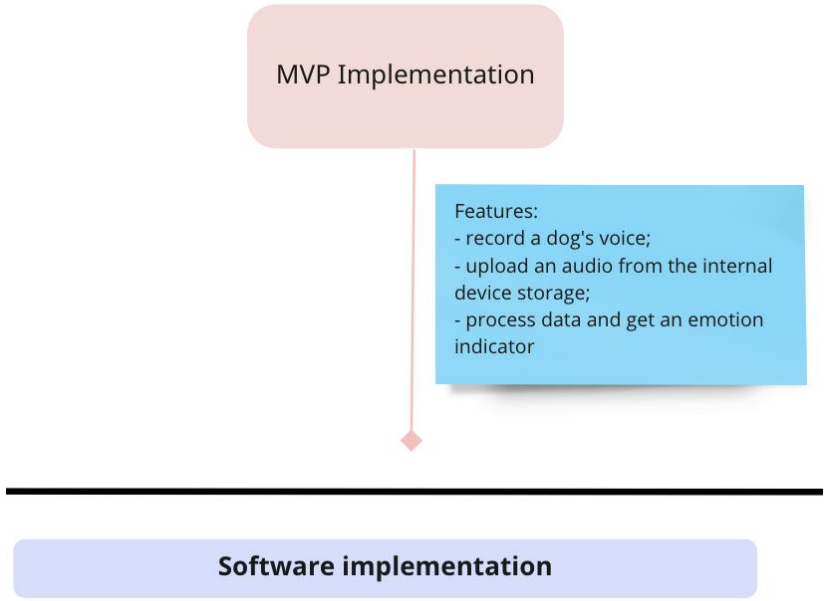
Sequence diagram II

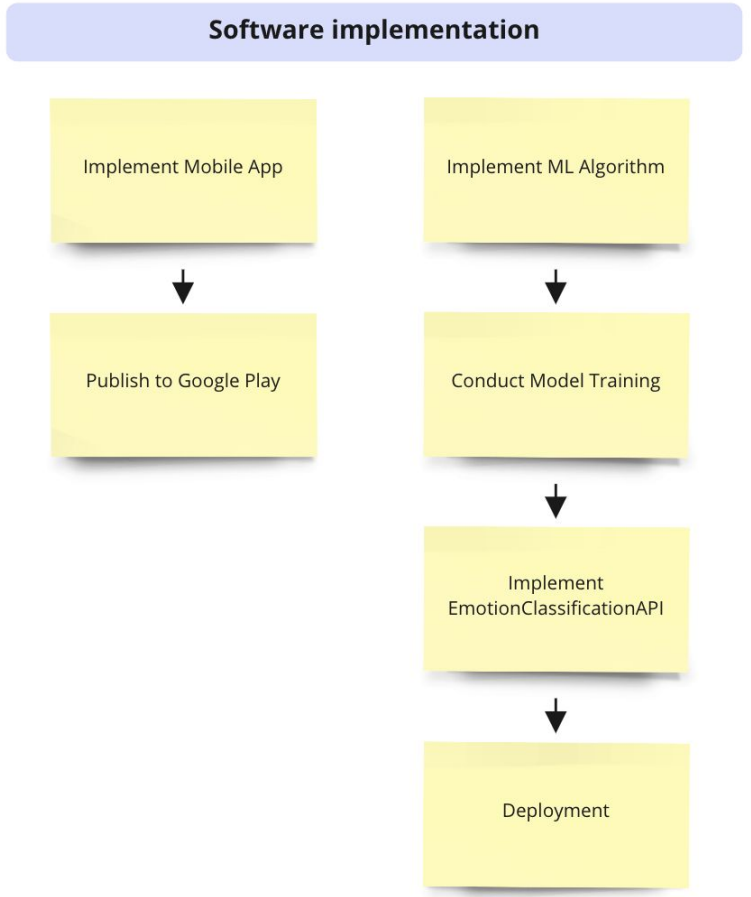


Development model

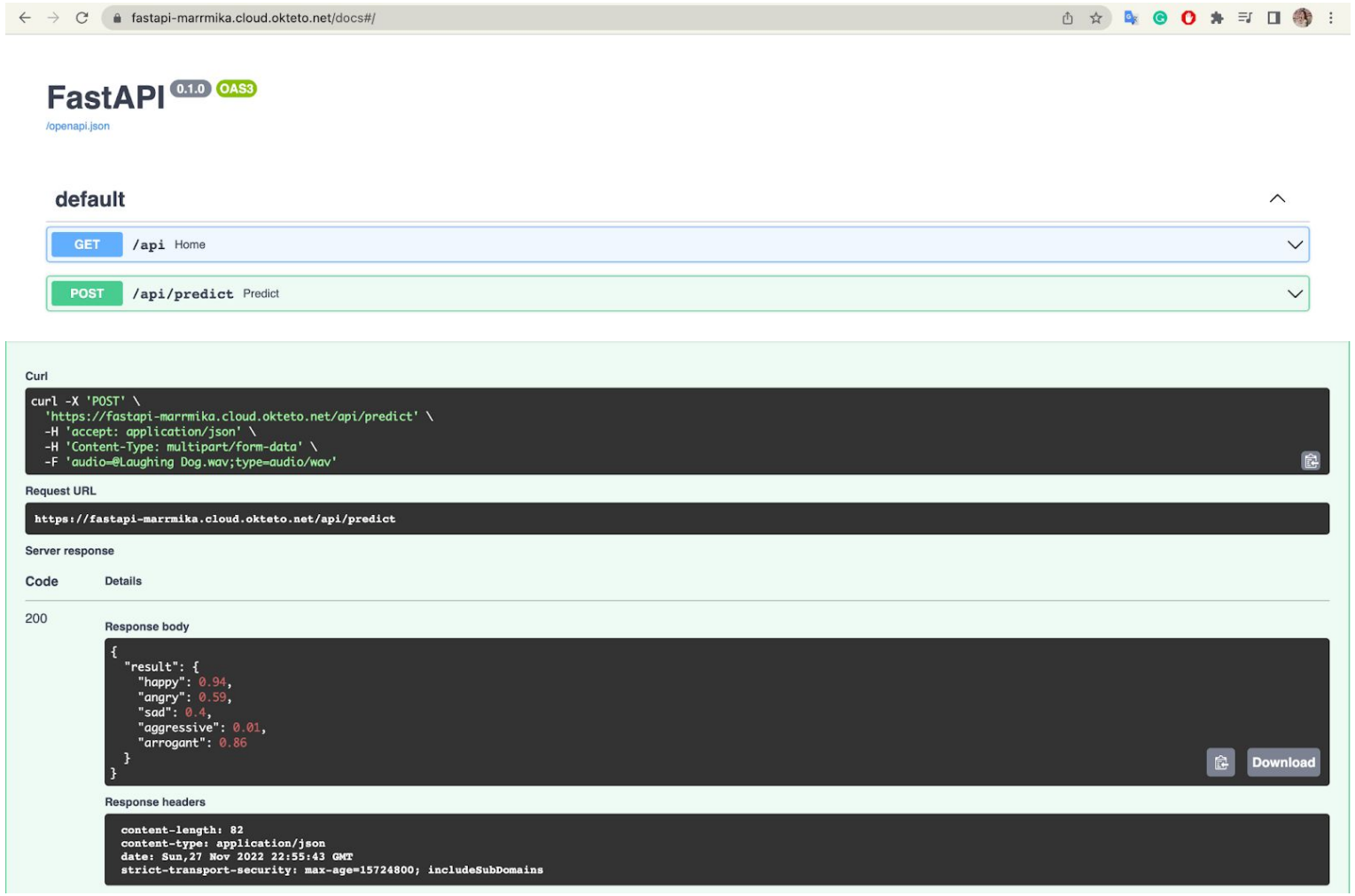


MVP

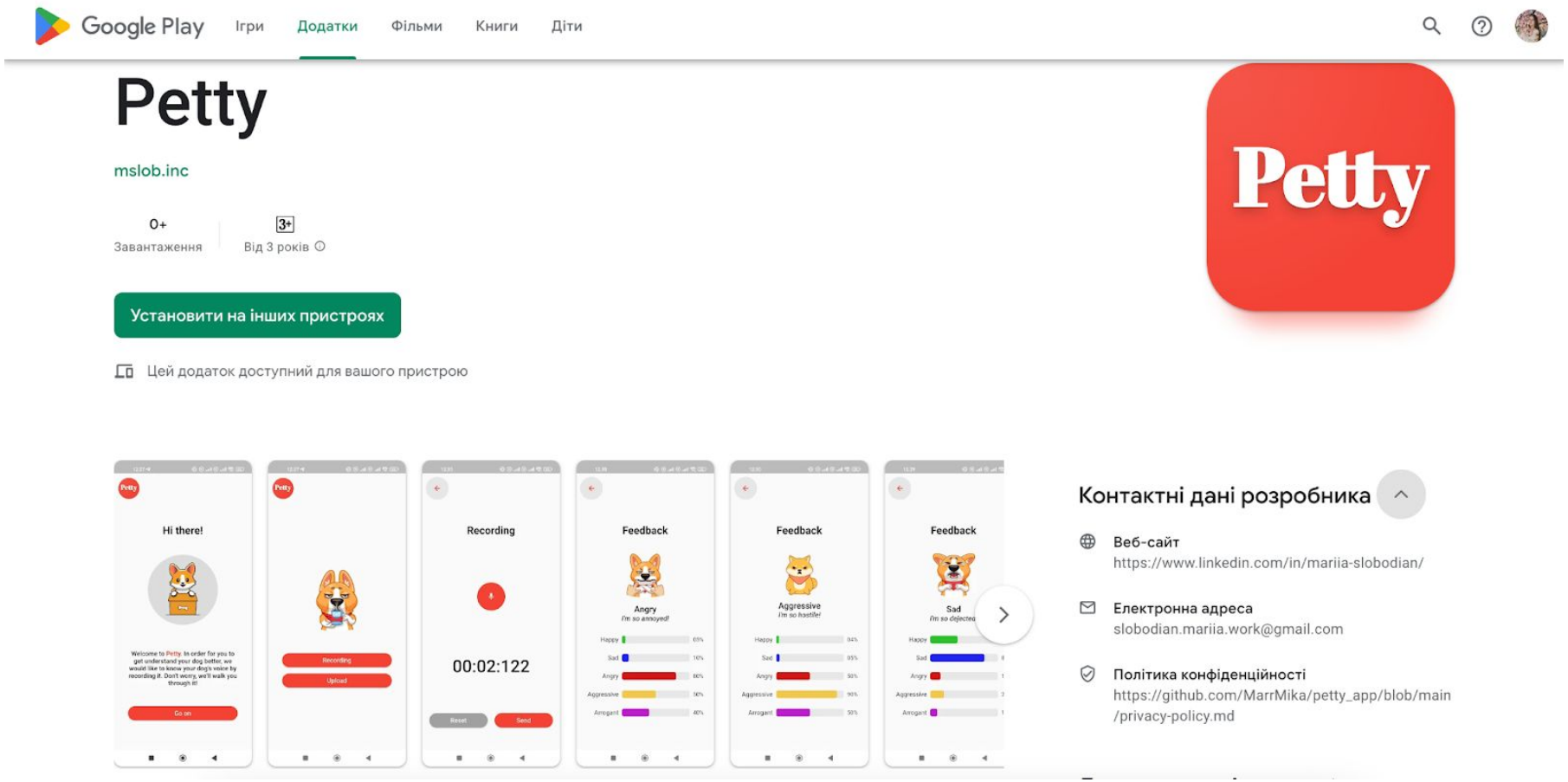

MVP Implementation
Features:
- record a dog's voice;
- upload an audio from the internal device storage;
- process data and get an emotion indicator
Software implementation


Software implementation
Implement Mobile App
Publish to Google Play
Implement ML Algorithm
Conduct Model Training
Implement EmotionClassificationAPI
Deployment

API Client


fastapi-marrmika.cloud.okteto.net/docs#/
FastAPI
0.1.0
OAS3
/openapi.json
default
GET /api Home
POST /api/predict Predict
Curl
curl -X 'POST' \
'https://fastapi-marrmika.cloud.okteto.net/api/predict' \
-H 'accept: application/json' \
-H 'Content-Type: multipart/form-data' \
-F 'audio=@Laughing Dog.wav;type=audio/wav'
Request URL
https://fastapi-marrmika.cloud.okteto.net/api/predict
Server response
Code
Details
200
Response body
{
"result": {
"happy": 0.94,
"angry": 0.59,
"sad": 0.4,
"aggressive": 0.01,
"arrogant": 0.86
}
}
Download
Response headers
content-length: 82
content-type: application/json
date: Sun,27 Nov 2022 22:55:43 GMT
strict-transport-security: max-age=15724800; includeSubDomains

Mobile App


Google Play
Ігри
Додатки
Фільми
Книги
Діти
Petty
mslob.inc
0+
Завантаження
3+
Від 3 років
Установити на інших пристроях
Цей додаток доступний для вашого пристрою
Petty
Hi there!
Recording
00:02:122
Feedback
Angry
Aggressive
Sad
Контактні дані розробника
Веб-сайт
https://www.linkedin.com/in/mariia-slobodian/
Електронна адреса
slobodian.mariia.work@gmail.com
Політика конфіденційності
https://github.com/MarrMika/petty_app/blob/main/privacy-policy.md

Results

- The iterative process of software development is described
- A mobile application and an application software interface for interaction with the method of classifying the emotional state of dogs have been developed
- The accuracy of the model has been improved up to 72% which is sufficient for the given task.

Discussion

- Despite the limited data, the ”fair and pain” category of barks continues to be the most recognizable.
- The lack of datasets for dogs’ sounds specifically makes it hard to train large complex models.
- Different results of applying the same model for the whole group of dogs and for each dog separately.

Thanks for your attention!