Speech Emotion Detection using IoT based Deep Learning

CSE496 – Graduation Project
Final Presentation

Ömer Faruk Bitikçioğlu

Advisor: Prof.Dr. Yusuf Sinan Akgül
January 2023

Content

- Project Diagram and Description
- Project Design Plan
- Project Timeline
- Data Collection & Augmentation
- Feature Extraction
- Modelling
- Tests & Results
- Application
- Success Criteria
- Conclusions
- References

Project Diagram and Description

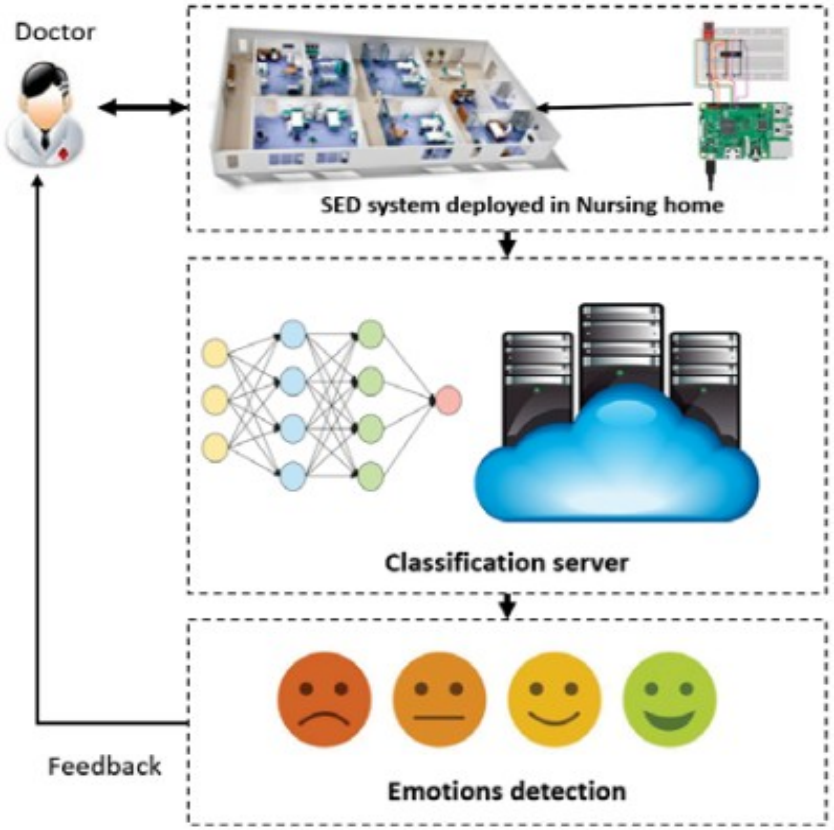


Fig. 1: The SED System: Real Time Speech Emotion Detection System based on Internet of Things (IoT) and Deep Learning for Health Care [1]

- Speech emotion detection system (SED)
- IoT device(mic) takes human voice as input
- Recognizes the human emotion
- Puts the recognition results on an interface
- Can be effective in many commercial applications

Project Design Plan

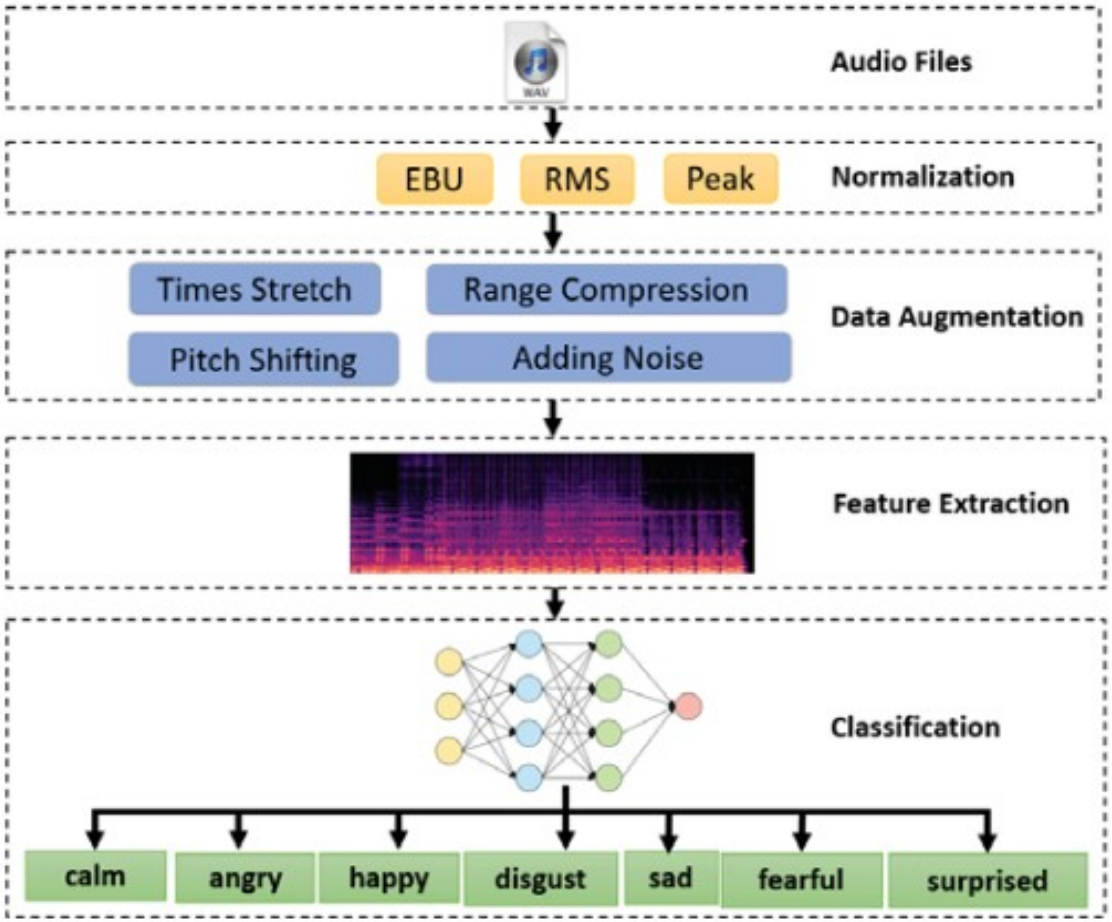


Fig. 2: The SED System: Classification Workflow for Speech Emotions [1]

- Take audio as input
- Normalization
- Augmentation
- Feature extraction
- Classification

Project Timeline

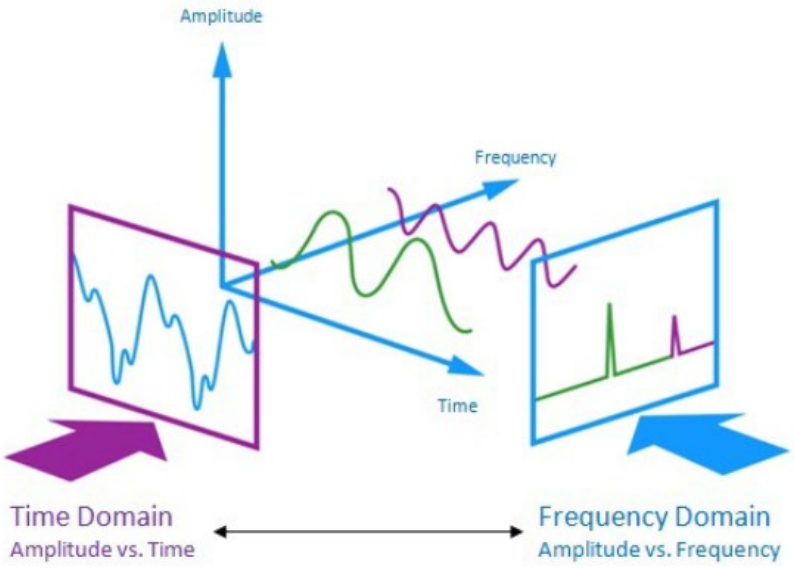
October 6	October 26	December 7	January 18	January 19
Collecting Turkish emotional speech data and labeling them into corresponding emotions.	Normalization, data augmentation, feature extraction, model training, classification, testing.	Development of speech application.	Preparing for the demo.	Writing report, preparing video.
START	1st Meeting	2nd Meeting	3rd Meeting	Demo

Data Collection & Augmentation

- Turkish emotion voice database (TurEV-DB) [2]
  - Angry(487), Calm(408), Happy(357), Sad(483), total 1735 samples.
- Data collected from variable sources.
  - Radio theaters
  - Series/movies
  - Podcasts
- Data Augmentation
- In total we have 5418 Turkish
  - emotional sample.

Feature Extraction

- Important step
- A three-dimensional signal
- Sample rate and sample data



In this project I am only extracting 5 features:

- Zero Crossing Rate
- Chroma_stft
- MFCC
- RMS(Root Mean Square) value
- MelSpectogram

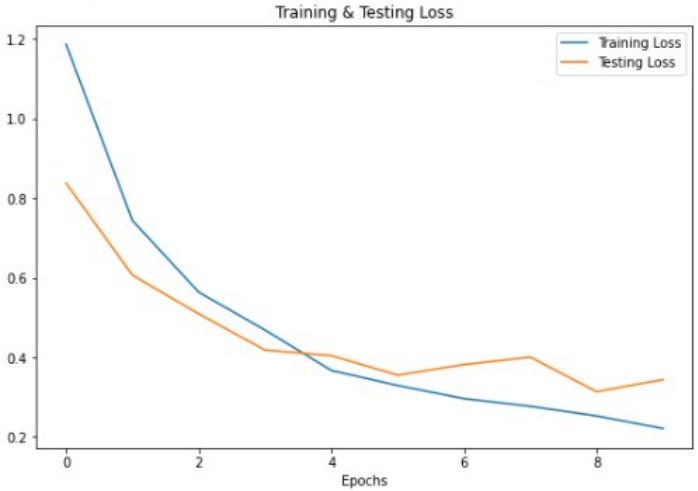
Modelling

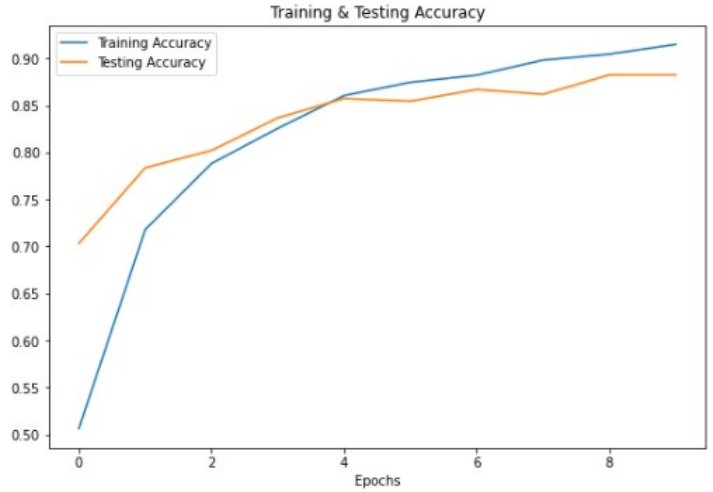
- CNN architecture
- MaxPooling
- Activation functions
  - ReLU
  - Softmax
- Adam optimizer

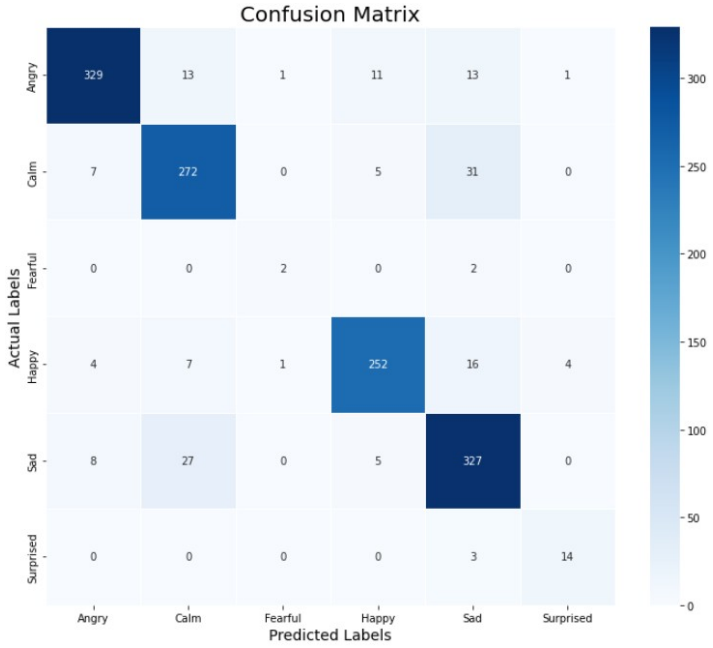
```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv1d (Conv1D)              (None, 162, 256)          1536
_________________________________________________________________
conv1d_1 (Conv1D)            (None, 162, 256)          327936
_________________________________________________________________
conv1d_2 (Conv1D)            (None, 162, 128)          163968
_________________________________________________________________
dropout (Dropout)            (None, 162, 128)          0
_________________________________________________________________
conv1d_3 (Conv1D)            (None, 162, 64)           41024
_________________________________________________________________
flatten (Flatten)            (None, 10368)             0
_________________________________________________________________
dense (Dense)                (None, 32)                331808
_________________________________________________________________
dropout_1 (Dropout)          (None, 32)                0
_________________________________________________________________
dense_1 (Dense)              (None, 6)                 198
=================================================================
Total params: 866,470
Trainable params: 866,470
Non-trainable params: 0
```

Tests & Results

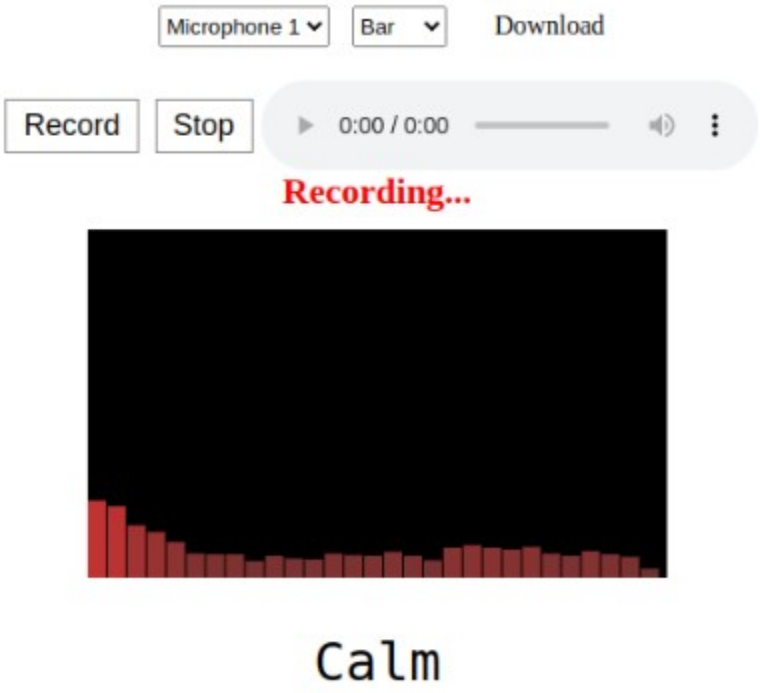
```
Accuracy of our model on test data :  88.26568126678467 %
```

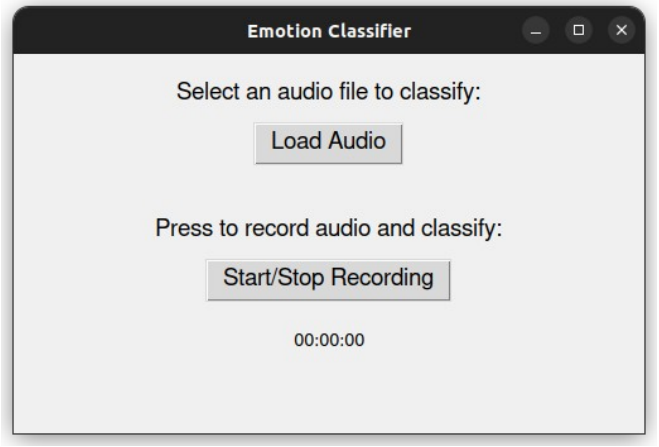


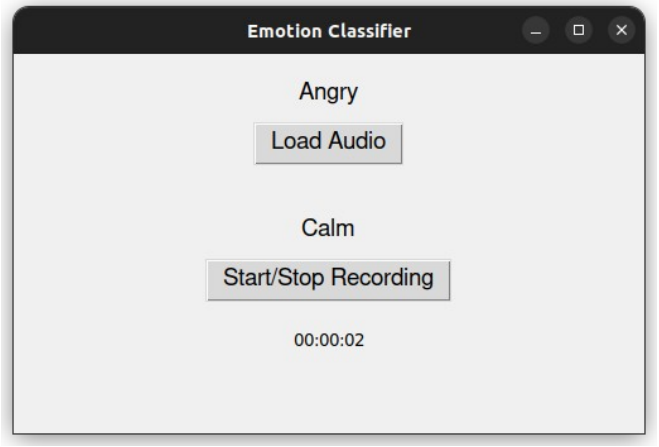




Application







Success Criteria

1) Turkish emotion detection with at least a 90% accuracy rate X - 88% accuracy

2)Emotion detection in real-time ✓

3)At least ten thousand data will be used (Collected data + augmented data) X – 5418

4)The model is capable of detecting seven different emotions (calm, happy, sad, angry, fearful, surprised, and disgusted) X – Six, no disgusted

Conclusions

- Successfully developed a deep learning model for speech emotion detection
- Achieved 88% average accuracy
- Sample applications are developed
- Can be applied in call centers and in-car systems
- Limitations: Lack of data and samples
- Further improvements and data collection can increase accuracy
- A promising approach for understanding emotions in speech.

References

1. Tariq, Z., Shah, S.K. and Lee, Y., 2019, December. Speech emotion detection using iot based deep learning for health care. In 2019 IEEE International Conference on Big Data (Big Data) (pp. 4191-4196). IEEE.
2. S. F. Canpolat, Z. Ormanoğlu, and D. Zeyrek, “Turkish emotion voice database (turev-db),” in Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Underresourced languages (SLTU) and Collaboration and Computing for UnderResourced Languages (CCURL), 2020, pp. 368–375.
3. N. Davis and K. Suresh, “Environmental sound classification using deep convolutional neural networks and data augmentation,” in 2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS). IEEE, 2018, pp. 41–45.
4. Latif, S., Rana, R., Younis, S., Qadir, J. and Epps, J., 2018. Transfer learning for improving speech emotion classification accuracy. arXiv preprint arXiv:1801.06353.