Speech Emotion Detection using IoT based Deep Learning

CSE496 – Graduation Project
2nd Meeting Presentation

Ömer Faruk Bitikçioğlu

Advisor: Prof.Dr. Yusuf Sinan Akgül
December 2022

Content

- Project Diagram and Description
- Project Design Plan
- Project Requirements
- Project Timeline
- What have we done?
- What we will do?
- Success Criteria
- References

Project Diagram and Description

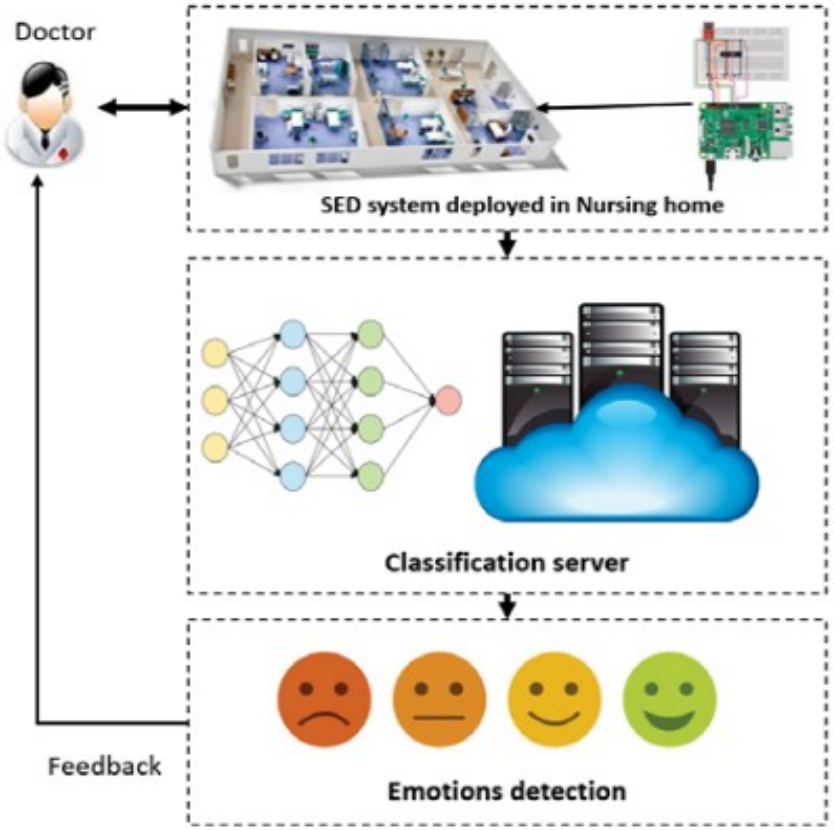


Fig. 1: The SED System: Real Time Speech Emotion Detection System based on Internet of Things (IoT) and Deep Learning for Health Care [1]

- Speech emotion detection system (SED)
- IoT device 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

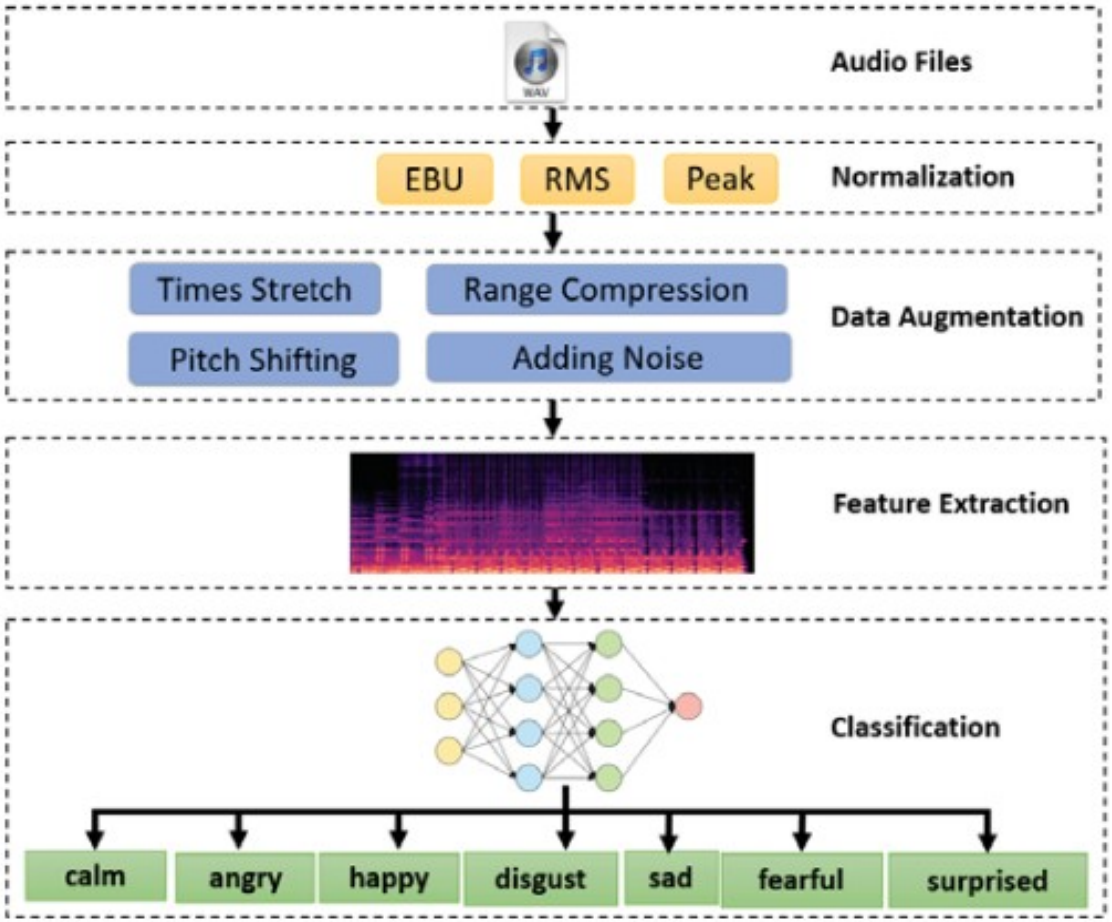
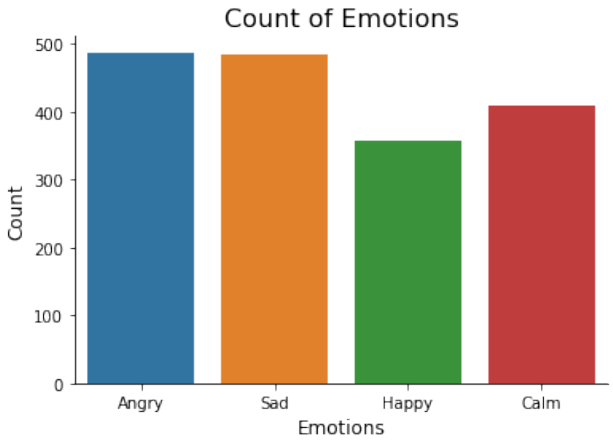


Fig. 2: The SED System: Classification Workflow for Speech Emotions [1]

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

Project Requirements

- Speech data from different speakers with different emotions ✓

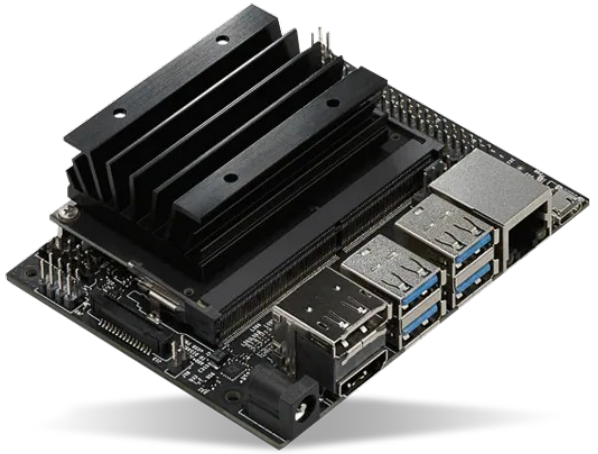


- Python libraries ✓

Ffmpeg, Librosa, numpy, sklearn, keras, pandas...

- Python for developing the speech application ✓

- Jetson Nano for IoT device ✓



- Microphone compatible with Jetson Nano ✓(USB mic)

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. Integration to Jetson Nano.	Integration tests. Preparing for the demo.	Writing report, preparing video.
START	1st Meeting	2nd Meeting	3rd Meeting	Demo

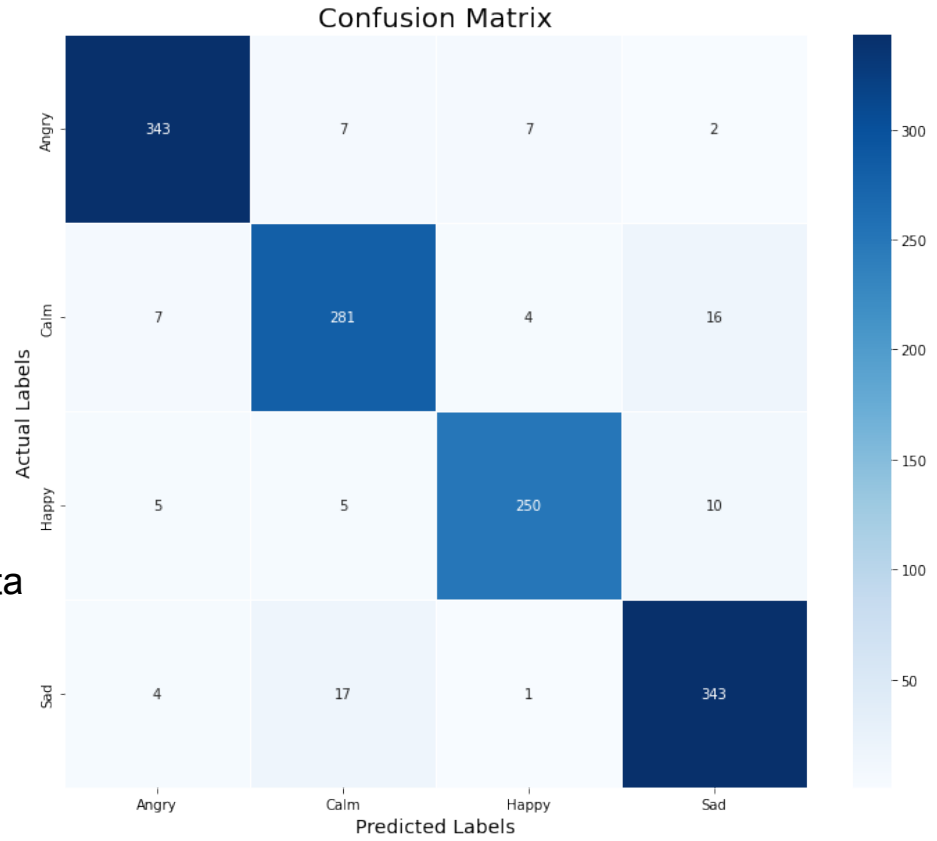
What have we done?

- Acquire Turkish emotional speech data ✓
  - 1734 data in total – Angry, Calm, Happy, Sad
- Use normalization techniques ✓
- Use data augmentation techniques ✓
  - Noise Injection, Stretching, Shifting, Pitch
- Extract features ✓
  - ZCR, Chroma_stft, MFCC, RMS, MelSpectogram
- Train the model ✓
- Use the model to classify speech data ✓

Success Criteria

1) Turkish emotion detection with at least a 90% accuracy rate ✓ (93% accuracy on standard data)

	precision	recall	f1-score	support
Angry	0.96	0.96	0.96	359
Calm	0.91	0.91	0.91	308
Happy	0.95	0.93	0.94	270
Sad	0.92	0.94	0.93	365
accuracy			0.93	1302
macro avg	0.94	0.93	0.93	1302
weighted avg	0.93	0.93	0.93	1302



2) Emotion detection in real-time (Prediction is about 0.2ms but overall test is not ready yet)

3) At least ten thousand data will be used (Collected data + augmented data) (5205 now)

4) The model is capable of detecting seven different emotions (calm, happy, sad, angry, fearful, surprised, and disgusted) (Only four now)

What we will do?

- Develop a speech application listens audio, classify it using the model, and show the result on the graphical interface
- Integration to Jetson Nano
- Collecting more Turkish data & improving model accuracy on different data sets
- Transfer learning techniques for the other emotions

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. L. R. Aguiar, M. Y. Costa, and N. C. Silla, “Exploring data augmentation to improve music genre classification with convnets,” in International Joint Conference on Neural Networks. IEEE, 2018, pp. 1–8.
3. S. Wei, K. Xu, D. Wang, F. Liao, H. Wang, and Q. Kong, “Sample mixed-based data augmentation for domestic audio tagging,” arXiv preprint arXiv:1808.03883, 2018.
4. 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.
5. B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, “librosa: Audio and music signal analysis in python,” in Proceedings of the 14th python in science conference, 2015, pp. 18–25.