Speech Emotion Recognition using Deep Learning

CSE496 – Graduation Project
Preliminary Presentation

Ömer Faruk Bitikçioğlu

Advisor: Prof.Dr. Yusuf Sinan Akgül
October 2022

Content

- Project Diagram and Description
- Project Design Plan
- Project Requirements
- Success Criteria
- References

Project Diagram and Description

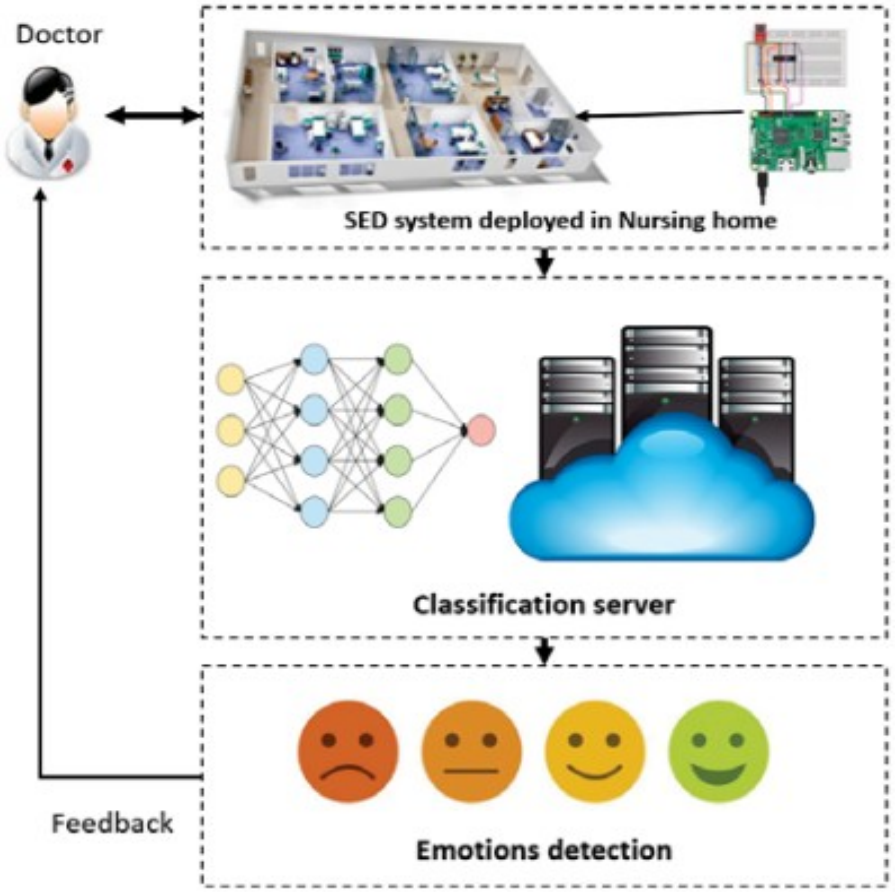

Fig. 1: The SED System: Real Time Speech Emotion Detection System based on Internet of Things (IoT) and Deep Learning for Health Care [1]

- Speech emotion detection system (SED)

- 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

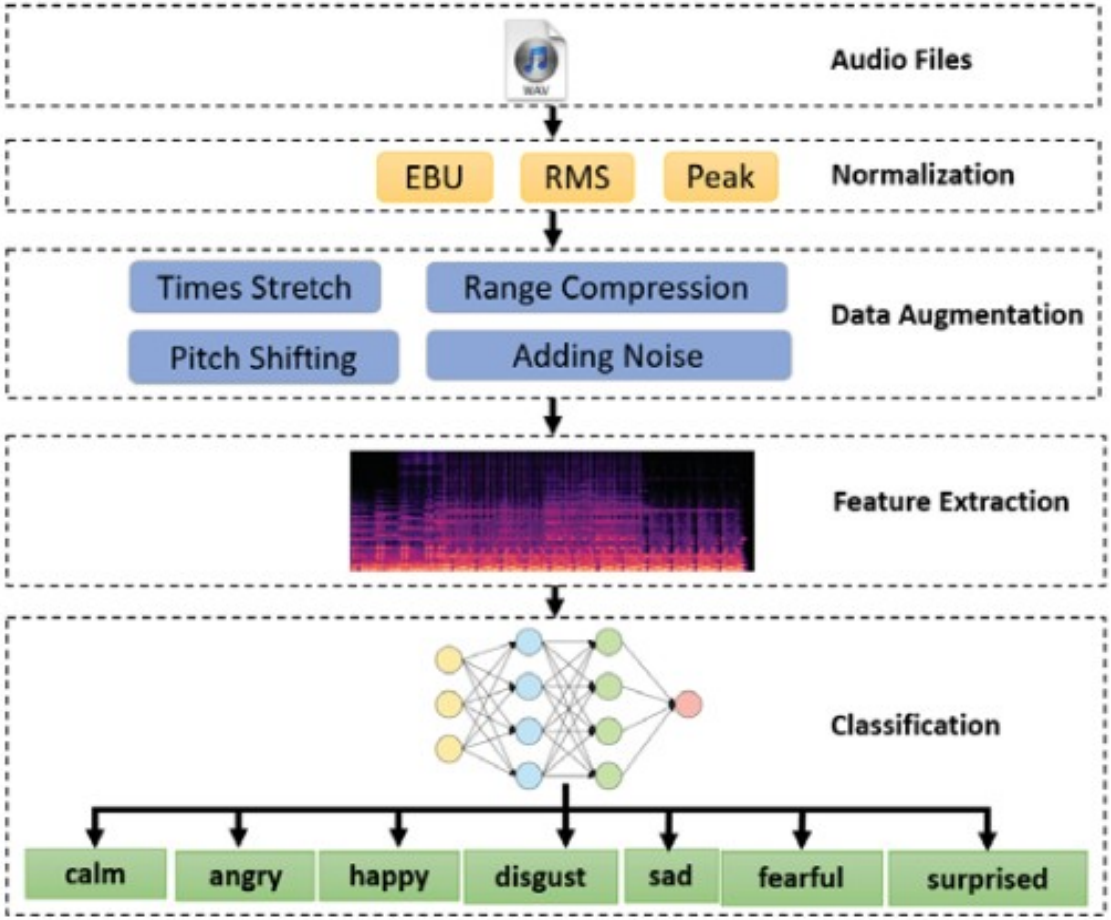

Fig. 2: The SED System: Classification Workflow for Speech Emotions [1]

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

Project Requirements - 1

What do we need to do to make this project work?

- Acquire emotional speech data from different speakers with different emotions.
- Use normalization techniques for the collected sound files. (EBU, RMS, Peak)
- Use data augmentation techniques to produce more data with the collected ones.
- Extract features from the data set.
- Use a deep learning model to classify sounds into different emotions.
- Develop an interface that can get emotion detection results and print nicely for demonstration purposes.
- Develop a speech server to get sound input, classify it and send the classification result to the interface developed earlier.
- Use an IoT device (Rasperry pi etc.) and make it possible to get sound input and send it to the speech server developed earlier.

Project Requirements - 2

- Speech data from different speakers with different emotions
- Python libraries
  - Ffmpeg, Librosa, numpy, sklearn, keras
- NodeJS (or similar) for developing the speech server
- Angular (or similar) for developing a user interface
- Raspberry Pi (or similar) for IoT communication

Success Criteria

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

2) Emotion detection in real-time (less than 1 second)

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

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

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.