

# Emotion Detection in Speech

by

Milad Shirani

# Contents

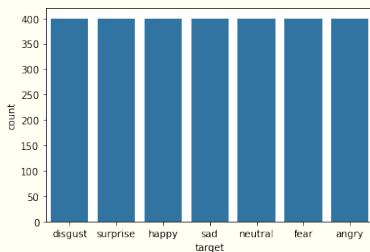
- ❏ Project Overview
- ❏ Data
- ❏ Modeling and Results
- ❏ Q & A

# Project Overview

1. In order for a therapist to fully analyze a patient, not only are the sentences important, but also the emotion conveyed by speech is important.
2. In this work, we want to introduce a new model to detect the emotion of an audio file.
3. We used several categorical models such as Logistic Regression, Decision Tree, Random Forest, XGBoost, LightGBM, as well as Convolutional Neural Network and Transfer Learning such as [EfficientNetB3](#) and [EfficientNetB7](#)

# Data and Method

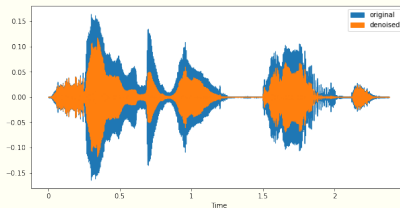
- 2800 audio files provided by [University of Toronto](#)



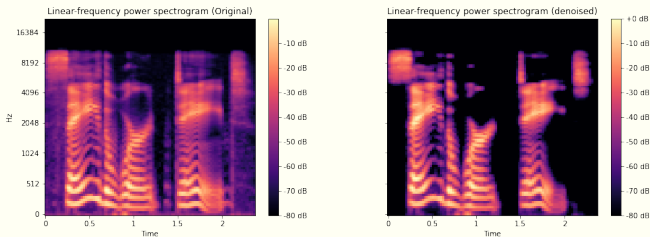
Distribution of Categories

- Converting audio files to mel-spectrograms

# Effects of Denoising

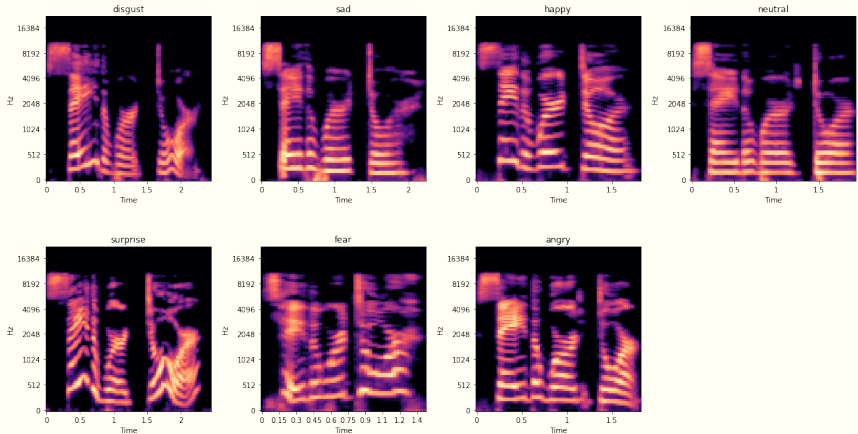


The Original and Denoised values of an Audio File



Mel-Spectrograms of Original and Denoised of an Audio File

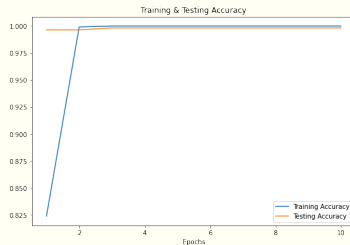
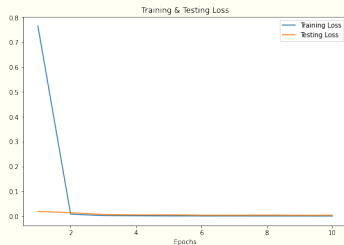
# Effects of emotion



Effects of emotion in saying the word "door"

# Modeling and Results

1. All the deep learning models performed well with test accuracy of 99% (except EfficientNetB7 which has the lowest test accuracy which is about 0.95% after 35 epochs).
2. We would recommend the first CNN model ([link to the model](#)) because it has the simplest structure and converges after 2 epochs as shown below



# Conclusion

1. The information in speech is conveyed through words and emotion.
2. Depending on how one pronounces a word, we can understand different meaning.
3. In order for a therapist to fully analyze a patient, it is important to understand both words and the emotion of the speech delivered by the patient.
4. The final model we introduce has the simplest structure.  
([link to the model](#))
5. This model can be implemented by virtual assistant such as Amazon Alexa or Siri as well in addition to the therapists.



# Next Steps

1. Gathering more data points for training.
2. Deploying neural networks by using LSTM or Conv1d layers and train them on numerical values obtained from audio files.
3. Trying using MFCCs (Mel Frequency Cepstral Coefficients) to train machine learning models.

*Thank  
You!*