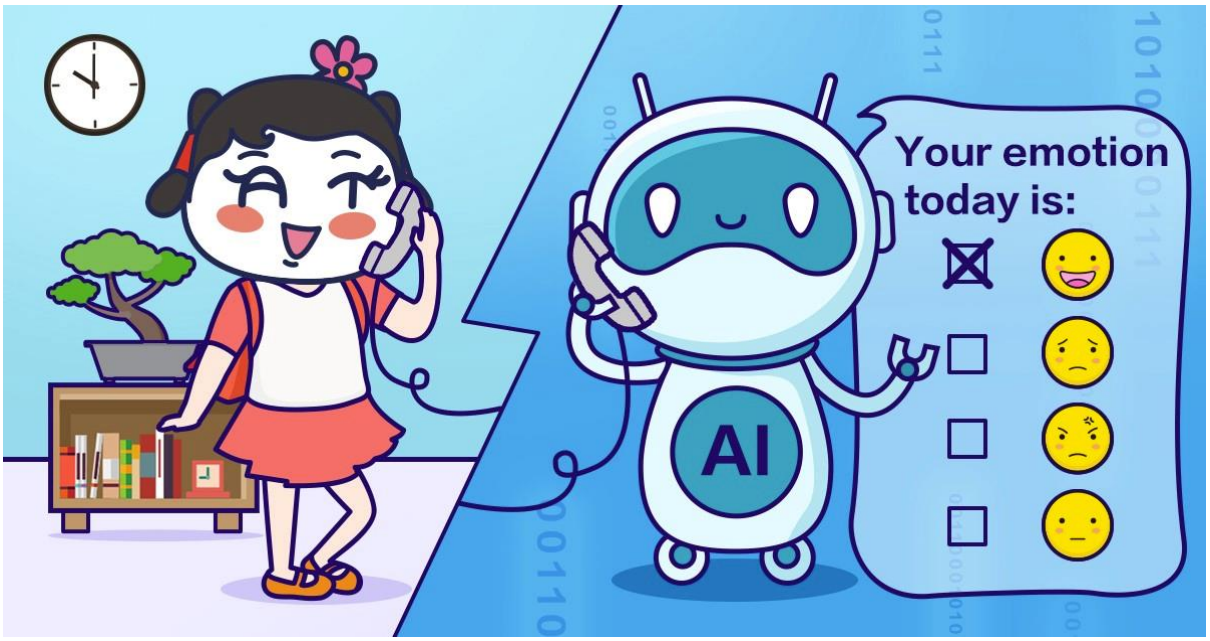
Emotions Classification Using Voices (Artificial Intelligence)

CSCE 5310 Methods in Empirical Analysis

Leela Varadatta Sai Addanki (11617487)
Sai Karthik Poojari (11518167)
Shashi Shekar Ranganathan (11533071)

Introduction:

Face detection and discrimination have always been innate abilities of humans. The same is now possible with computers. There are many new opportunities opened up by this. Face detection and recognition can be used to enhance access and security, as is the case with the most recent Apple iPhone (see gif below), to process payments devoid of physical cards — the iPhone also does this!, to enable criminal identification, and to offer individualized healthcare and other services. Face recognition and identification has received much research, and there are many internet tools for it. The exponential increase of Internet multimedia traffic has been facilitated by real-time multimedia applications and services such as video conferencing, telepresence, real-time content distribution, telemedicine, voice commands on wearables, and online gaming. Systems for acquiring, processing, and transmitting multimedia information are rich sources of integrated audio, text, and video streams. A vast amount of information is included in this massively increasing internet traffic of human talks, particularly voice-related characteristics that assist in describing human behavior and underlying emotions. People's thoughts and actions in day-to-day events are influenced by their emotions.



Importance:

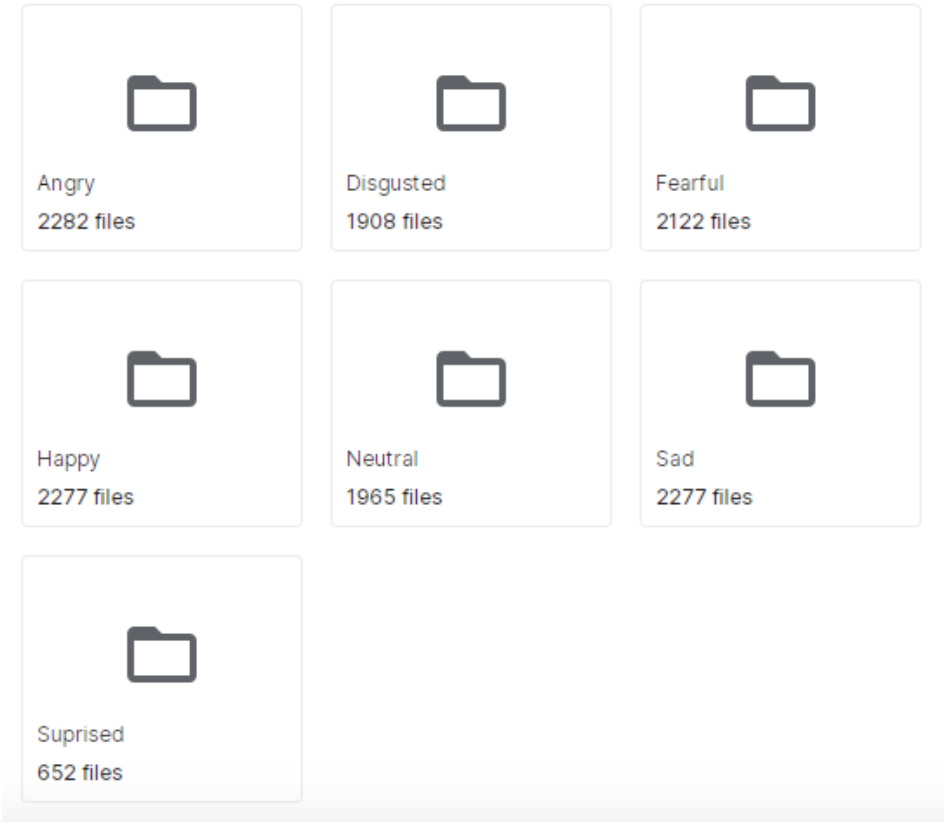
Human-to-human communication depends on affect recognition, and emotion recognition is a necessary component of a successful connection between people and technology. These days, a growing number of intelligent systems use emotion detection models to enhance their communication with people. Candidates' feelings during interviews and their responses to certain topics can be assessed using emotion recognition. The interview process may be made more efficient for prospective applicants by using the information below. Sentiment analysis has a broad use in a variety of fields, including biometric protection and social interface. In order to replicate the human brain, it employs a variety of both supervised and unsupervised machine-learning techniques, which gives us insight into artificial intelligence or machine intelligence. Emotions are significant drivers of future behavior because they get our body ready for quick action, have an impact on our ideas, and can be felt. A lot of us aspire to feel joy, pride, contentment, or victory in our successes and accomplishments. Therefore, it would be preferable if we could identify emotions in people and adjust our strategy accordingly.

Approach:

We used the LSTM model for this system building because LSTM are very good models for Sequential data. As our data is sequential so we have used LSTM. An artificial neural network called long short-term memory is utilized in deep learning and artificial intelligence. LSTM features feedback connections as opposed to typical feedforward neural networks. Such a recurrent neural network may analyze complete data sequences in addition to single data points. Because it enables the learning of even more parameters, the LSTM cell increases long-term memory in an even more efficient manner. As a result, it is the most effective [Recurrent Neural Network] for predicting, particularly when your data show a longer-term trend.

Dataset Used:

The dataset used for this model training and test contains 7 emotions data (Angry, Disgusted, Fearful, Happy, Neutral, Sad, Surprised) for every emotion there are almost 2000 audio files in the dataset.



Steps:

1. Collect the data for the model
2. Load the audio files and extract the features from it
3. Normalize the features of the audios
4. Build the LSTM model
5. Split the dataset of the audios
6. Train the model
7. Test the model
8. Evaluate the model

Related Work:

Computers can comprehend facial expressions based on human thought thanks to facial expression recognition, which is a vital component of human-computer interaction. According to the facial expression recognition process, there are three key modules: face detection, feature extraction, and classification. Face detection, which is considered the key technology of face recognition (Adjabi et al., 2020; Zhang et al., 2021), has reached basic maturity with its rapid development. It can effectively extract from the original face image of excellent characteristics, and the characteristics of correct classification become a key factor affecting facial expression recognition results. Gao and Ma (2020), for instance, extracted facial expression characteristics from facial photos in order to forecast emotional states based on changes in facial expression. By examining and evaluating the physical properties of the speaker's speech in various emotional states, the emotion identification technique based on speech is able to recognise and assess the speaker's current emotional information. The use of speech signals to emotion recognition by Ton-That and Cao (2019) produced positive outcomes on the voice emotion database. However, on the one hand, individual variations will result in significant variations in speech signals, necessitating the creation of a sizable phonetic database, which may provide some recognising challenges. On the other hand, a loud environment will degrade speech quality and make it harder to recognise emotions. For this reason, the acquisition of speech signals places a significant demand on the environment.

References:

1. Abeysinghe S et al. (2018) Enhancing decision making capacity in tourism domain using social media analytics. 2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer), pp 369–375. https://doi.org/10.1109/ICTER.2018.8615462
2. Adikari A, Alahakoon D (2021) Understanding citizens' emotional pulse in a smart city using artificial intelligence. IEEE Trans Ind Inf 17(4):2743–2751. https://doi.org/10.1109/TII.2020.3009277
3. Adikari A, Burnett D, Sedera D, de Silva D, Alahakoon D (2021) Value co-creation for open innovation: An evidence-based study of the data driven paradigm of social media using machine learning. Int J Inf Manag Data Insights 1(2):100022
4. Adikari A, Nawaratne R, De Silva D, Ranasinghe S, Alahakoon O, Alahakoon D (2021) Emotions of COVID-19: Content analysis of self-reported information using artificial intelligence. J Med Internet Res 23(4):e27341
5. Adikari A, Gamage G, de Silva D, Mills N, Wong S, Alahakoon D (2021) A self structuring artificial intelligence framework for deep emotions modeling and analysis on the social web. Futur Gener Comput Syst 116:302–315

   Github URL: https://github.com/Shashisekharranganathan/Emotion-Classification-using-Audio