International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

SPEECH EMOTION RECOGNITION

Chintala Santoshi Reddy[1] , Nandini Jena[2], Pamireddy Meghana[3]

Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women, Telangana, India

Abstract

Speech Emotion Recognition, abbreviated as SER, is the act of attempting to recognize human emotion and affective states from speech. This is capitalizing on the fact that voice often reflects underlying emotion through tone and pitch. This is also the phenomenon that animals like dogs and horses employ to be able to understand human emotion. SER is tough because emotions are subjective and annotating audio is challenging.The main objective of this SER, recognizing the emotional state of the speaker from his/her speaker. Various emotions like happiness, sadness, anger, etc. can be recognized. SER uses speech processing to recognize the emotional state. Speech Processing is one of the important branches of digital signal processing and finds applications in Human-computer interfaces, Telecommunication, Assistive technologies, Audio mining, Security, and so on. Speech Emotion Recognition is important to have a natural interaction between human beings and machines. In SER, the emotional state of a speaker is extracted from his/her speech. The acoustic characteristic of the speech signal is Feature. Feature extraction is the process that extracts a small amount of data from the speech signal that can later be used to represent each speaker. Many feature extraction methods are available and Mel Frequency Cepstral Coefficient (MFCC) is the commonly used method. In this project, speaker emotions are recognized using the data extracted from the speaker's voice signal. Mel Frequency Cepstral Coefficient (MFCC) technique is used to recognize the emotion of a speaker from their voice.

1. Introduction

1.1 About Project

Speech Emotion Recognition recognizes the emotional state of the speaker from his/her speaker. Various emotions like happiness, sadness, anger, etc, can be recognized. SER uses speech processing to recognize the emotional state. Speech Processing is one of the important branches of digital signal processing and finds applications in Human-computer interfaces, Telecommunication, Assistive technologies, Audio mining, Security, and so on. Speech Emotion Recognition is important to have a natural interaction between human beings and machines. In SER, the emotional state of a speaker is extracted from his/her speech. The acoustic characteristic of the speech signal is Feature. Feature extraction is the process that extracts a small amount of data from the speech signal that can later be used to represent each speaker. Many feature extraction methods are available and Mel Frequency Cepstral Coefficient (MFCC) is the most commonly used method. In this project, speaker emotions are recognized using the data extracted from the speaker's voice signal. Mel Frequency Cepstral Coefficient (MFCC) technique is used to recognize the emotion of a speaker from their voice.

1.2 Objectives of the Project

We are working with four different datasets, so we will be creating a dataframe, Ravdess DF, Crema-D DF, Savee DF and Tess DF, storing all emotions of the data in dataframe with their paths. The aim of Speech Emotion Recognition is to train the code to give accuracy, in other words is to find the highest accuracy of the dataset.

Also, we tried to perform analysis with single and combinations of datasets, to obtain the highest accuracy from his/her speech. We will also be using the datasets together with different combinations and also used them individually.

1.3 Scope of the Project

Speech Emotion Recognition uses speech processing to recognize the emotional state. It recognizes the emotional state of the speaker from his/her speaker. This system uses four different datasets to find the highest providing accuracy dataset among them. As we are working with four different datasets, so we will be creating a dataframe storing all emotions of the data in dataframe with their paths. We will use this dataframe to extract features for our model training. For each dataset we have data frame. This is Ravdess DF, Crema-D DF, Savee DF and Tess DF.

2. Literature Survey

2.1 Existing System

In the existing system, The speech sample is first passed through the reference database which is maintained for recognition of emotion. For each frame, MFCC(Mel Frequency Cepstral Coefficient) is calculated as the main feature for emotion recognition. The reference database is maintained which contains the MFCC's of emotions that is happiness, sadness, anger, etc.

2.2 Proposed System

In the proposed work, a human voice is given as input based on them it trains itself and identifies if the given emotion as happy, sad, anger, disgust, etc. And also find which is the most effective dataset.

3. Proposed Architecture

First the audio signals are given into the model then the features are selected from the datasets. After selected and extracted the data is trained and the testing and training loss are found after that we displayed the confusion matrix and the classification report so that we can find which dataset is effective to use while doing speech emotion recognition related models.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

3.1 Architecture

The following is the architecture of this project.

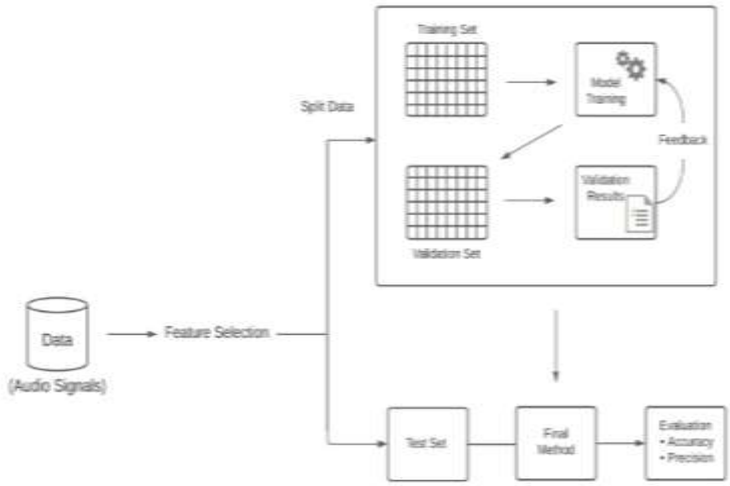


Fig3.1: Architecture

3.2 Block Diagram

. A block diagram is a diagram of a system in which the principal parts or functions are represented by blocks connected by lines that show the relationships of the blocks. They are heavily used in engineering in hardware design, electronic design, software design, and process flow diagrams.

In the below block diagram we can see the audio signals are fed into the model and model extracts the features we have selected mfcc, chroma, ZCR, shift in our model. Then we fed the data into the classifier. The classifier detects the emotions and displays it. But in our project we are just finding the accuracy so that it would be easy for the ones who want to create a model which detects the emotion accurately.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

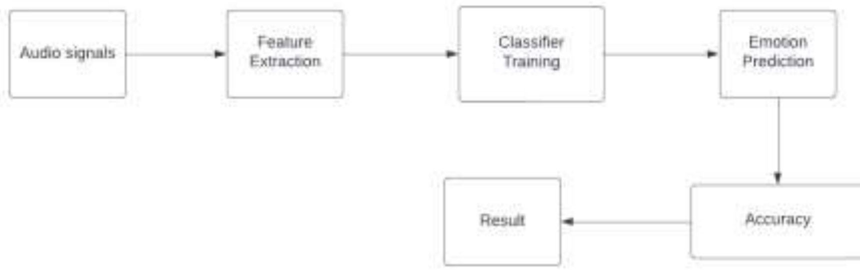


Fig3.2: Block Diagram

4. Implementation

4.1 Algorithm

Step1:Importing Libraries

We used pandas,numpy, os, sys, librosa, librosa.display, seaborn, matplotlib, sklearn.preprocessing, sklearn.metrics, sklearn.model_selection, IPython.display, keras in our project.

Step2: Data Preparation

We prepare the datasets for training the model. The datasets that we used in this project are Ravdess Crema Tess and Savee.

Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)

Speech audio-only files (16bit, 48kHz .wav) from the RAVDESS. Full dataset of speech and song, audio and video (24.8 GB) available from Zenodo. Construction and perceptual validation of the RAVDESS is described in our Open Access paper in PLoS ONE.
Files.

This portion of the RAVDESS that we used contains 1440 files: 60 trials per actor x 24 actors = 1440. The RAVDESS contains 24 professional actors (12 female, 12 male), vocalizing two lexically-matched statements in a neutral North American accent. Speech emotions includes calm, happy, sad, angry, fearful, surprise, and disgust expressions. Each expression is produced at two levels of emotional intensity (normal, strong), with an additional neutral expression.

File naming convention

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Each of the 1440 files has a unique filename. The filename consists of a 7-part numerical identifier (e.g., 03-01-06-01-02-01-12.wav). These identifiers define the stimulus characteristics:

Filename identifiers

- Modality (01 = full-AV, 02 = video-only, 03 = audio-only).
- Vocal channel (01 = speech, 02 = song).
- Emotion (01 = neutral, 02 = calm, 03 = happy, 04 = sad, 05 = angry, 06 = fearful, 07 = disgust, 08 = surprised).
- Emotional intensity (01 = normal, 02 = strong). NOTE: There is no strong intensity for the 'neutral' emotion.
- Statement (01 = "Kids are talking by the door", 02 = "Dogs are sitting by the door").
- Repetition (01 = 1st repetition, 02 = 2nd repetition).
- Actor (01 to 24. Odd numbered actors are male, even numbered actors are female).

Filename example: 03-01-06-01-02-01-12.wav

1. Audio-only (03)
2. Speech (01)
3. Fearful (06)
4. Normal intensity (01)
5. Statement "dogs" (02)
6. 1st Repetition (01)
7. 12^{th} Actor (12) Female, as the actor ID number is even.

Toronto emotional speech set (TESS)

TESS dataset is a female only and is of very high quality audio. Most of the other dataset is skewed towards male speakers and thus brings about a slightly imbalance representation. So because of that, this dataset would serve a very good training dataset for the emotion classifier in terms of generalisation (not overfitting).

There are a set of 200 target words were spoken in the carrier phrase "Say the word _' by two actresses (aged 26 and 64 years) and recordings were made of the set portraying each of seven emotions (anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral). There are 2800 data points (audio files) in total.

The dataset is organised such that each of the two female actor and their emotions are contain within its own folder. And within that, all 200 target words audio file can be found. The format of the audio file is a WAV format

Crowd Sourced Emotional Multimodal Actors Dataset (CREMA-D)

CREMA-D dataset is the sheer variety of data which helps train a model that can be generalised across new datasets. Many audio datasets use a limited number of speakers which leads to a lot of information leakage. CREMA-D has many speakers.

CREMA-D is a data set of 7,442 original clips from 91 actors. These clips were from 48 male and 43 female actors between the ages of 20 and 74 coming from a variety of races and ethnicities (African America, Asian, Caucasian, Hispanic, and Unspecified). Actors spoke from a selection of 12 sentences. The sentences were presented using one of six different emotions (Anger, Disgust, Fear, Happy, Neutral, and Sad) and four different emotion levels (Low, Medium, High, and Unspecified).

Surrey Audio-Visual Expressed Emotion (SAVEE)

SAVEE dataset is male only and is of very high quality audio. Because the male only speaker will bring about a slightly imbalance representation, it would be advisable to complement other datasets with more female speakers.

The SAVEE database was recorded from four native English male speakers (identified as DC, JE, JK, KL), postgraduate students and researchers at the University of Surrey aged from 27 to 31 years. Emotion has been described psychologically in discrete categories: anger, disgust, fear, happiness, sadness and surprise. A neutral category is also added to provide recordings of 7 emotion categories.

The text material consisted of 15 TIMIT sentences per emotion: 3 common, 2 emotion-specific and 10 generic sentences that were different for each emotion and phonetically-balanced. The 3 common and 2 × 6 = 12 emotion-specific sentences were recorded as neutral to give 30 neutral sentences. This resulted in a total of 120 utterances per speaker.

Step3: Data Visualisation and Exploration

Waveplot:

Waveplots let us know the loudness of the audio at a given time.
To visualize sound we can use two of these modules

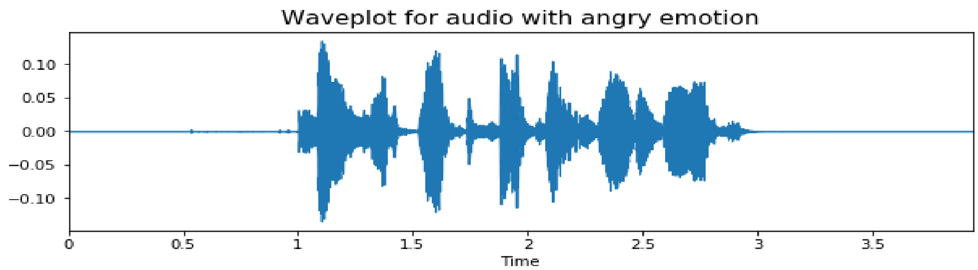
- Matplotlib
- Numpy

Approach

- Import matplotlib, Numpy, wave, and sys module.
- Open the audio file using the *wave.open()* method.
- Read all frames of the opened sound wave using *readframes*() function.
- Store the frame rate in a variable using the *getframrate*() function.
- Finally, plot the x-axis in seconds using frame rate.
- Use the matplotlib.figure() function to plot the derived graph

- Use labels as per the requirement.

The below is one of the wave plotted for the emotion anger.

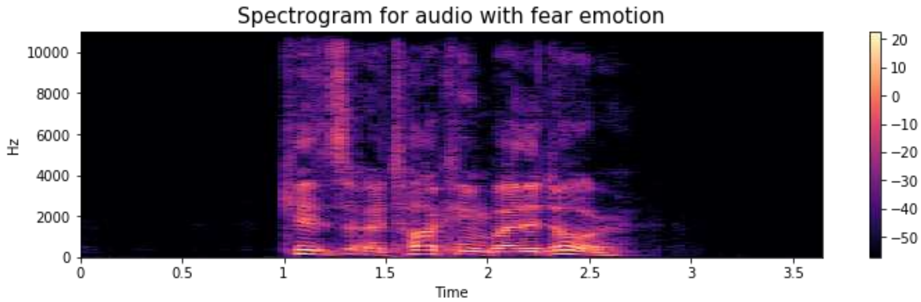


Spectrogram:

A spectrogram is a visual way of representing the signal strength, or "loudness", of a signal over time at various frequencies present in a particular waveform. Not only can one see whether there is more or less energy at, for example, 2 Hz vs 10 Hz, but one can also see how energy levels vary over time. In other sciences spectrograms are commonly used to display frequencies of sound waves produced by humans, machinery, animals, whales, jets, etc., as recorded by microphones. In the seismic world, spectrograms are increasingly being used to look at frequency content of continuous signals recorded by individual or groups of seismometers to help distinguish and characterize different types of earthquakes or other vibrations in the earth.

Spectrograms are basically two-dimensional graphs, with a third dimension represented by colors. Time runs from left (oldest) to right (youngest) along the horizontal axis. Each of our volcano and earthquake sub-groups of spectrograms shows 10 minutes of data with the tic marks along the horizontal axis corresponding to 1-minute intervals. The vertical axis represents frequency, which can also be thought of as pitch or tone, with the lowest frequencies at the bottom and the highest frequencies at the top. The amplitude (or energy or "loudness") of a particular frequency at a particular time is represented by the third dimension, color, with dark blues corresponding to low amplitudes and brighter colors up through red corresponding to progressively stronger (or louder) amplitudes.

The below is an example for the spectrogram.



Step 4: Data Augmentation

- Data augmentation is the process by which we create new synthetic data samples by adding small perturbations on our initial training set.
- To generate syntactic data for audio, we can apply noise injection, shifting time, changing pitch and speed.
- The objective is to make our model invariant to those perturbations and enhace its ability to generalize.
- In order to this to work adding the perturbations must conserve the same label as the original training sample.
- In images data augmentation can be performed by shifting the image, zooming, rotating
- We used noise, stretching (ie. changing speed) and some pitching.

Step 5: Feature Extraction

In speaker independent speech recognition, a premium is placed on extracting features that are somewhat invariant to changes in the speaker. So feature extraction involves analysis of speech signal. Broadly the feature extraction techniques are classified as temporal analysis and spectral analysis technique. In temporal analysis the speech wave form itself is used for analysis. In spectral analysis spectral representation of speech signal is used for analysis.

Types of transformation that one can perform are:

1. Zero Crossing Rate : The rate of sign-changes of the signal during the duration of a particular frame.
2. Energy : The sum of squares of the signal values, normalized by the respective frame length.
3. Entropy of Energy : The entropy of sub-frames' normalized energies. It can be interpreted as a measure of abrupt changes.
4. Spectral Centroid : The center of gravity of the spectrum.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

5. Spectral Spread : The second central moment of the spectrum.
6. Spectral Entropy : Entropy of the normalized spectral energies for a set of sub-frames.
7. Spectral Flux : The squared difference between the normalized magnitudes of the spectra of the two successive frames.
8. Spectral Roll off : The frequency below which 90% of the magnitude distribution of the spectrum is concentrated.
9. MFCCs Mel Frequency Cepstral Coefficients form a cepstral representation where the frequency bands are not linear but distributed according to the mel-scale.
10. Chroma Vector : A 12-element representation of the spectral energy where the bins represent the 12 equal-tempered pitch classes of western-type music (semitone spacing).
11. Chroma Deviation : The standard deviation of the 12 chroma coefficients.

Step 6:Data Preparation(Normalization)

Normalization is a technique often applied as part of data preparation for machine learning. The goal of normalization is to change the values of numeric columns in the dataset to use a common scale, without distorting differences in the ranges of values or losing information. Normalization is also required for some algorithms to model the data correctly.

Step 7: Modelling

We use 14 layers in our model. The below is the model summary of the model in which we used all the datasets

Layer (type)	Output Shape	Param #
conv1d_28 (Conv1D)	(None, 162, 256)	1536
max_pooling1d_28 (MaxPooling	(None, 81, 256)	0
conv1d_29 (Conv1D)	(None, 81, 256)	327936
max_pooling1d_29 (MaxPooling	(None, 41, 256)	0
conv1d_30 (Conv1D)	(None, 41, 128)	163968
max_pooling1d_30 (MaxPooling	(None, 21, 128)	0
dropout_13 (Dropout)	(None, 21, 128)	0
conv1d_31 (Conv1D)	(None, 21, 64)	41024
max_pooling1d_31 (MaxPooling	(None, 11, 64)	0

flatten_7 (Flatten)	(None, 704)	0
dense_13 (Dense)	(None, 32)	22560
dropout_14 (Dropout)	(None, 32)	0
dense_14 (Dense)	(None, 8)	264

===

Step 8: Prediction and Accuracy

The emotions are detected and after this we will also find accuracy of the model by its prediction.

It is the process of automatically choosing relevant features for your machine learning model based on the type of problem you are trying to solve. We do this by including or excluding important features without changing them. It helps in cutting down the noise in our data and reducing the size of our input data.

Machine learning model accuracy is the measurement used to determine which model is best at identifying relationships and patterns between variables in a dataset based on the input, or training, data. The better a model can generalize to 'unseen' data, the better predictions and insights it can produce, which in turn deliver more business value.

Step 9: Result

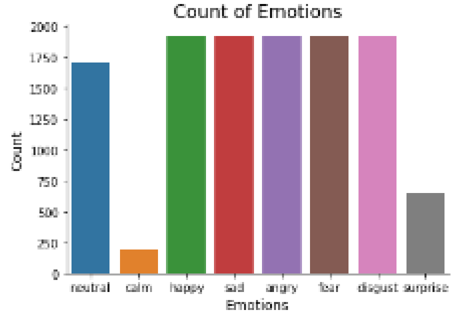
In the result we have shown the graph about the losses, the training and testing loss. And the also displayed the table which has all the precision, recall, f1-score, support.testing .We will use ML algorithms to categorize input into four categories .And we also calculates metrics like accuracy,f1 measure, recall, precision.

4.2 Code Implementation

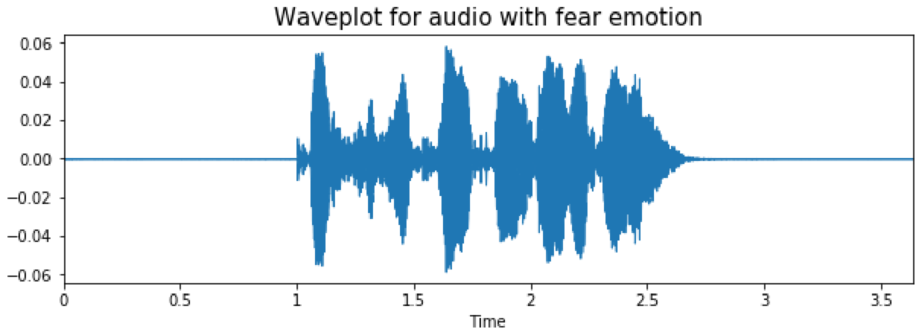
RAVDESS+CREMA+SAVEE+TESS

The below bar graph shows the count of emotions. Neutral is between 1700-1750, whereas calm is between 200-250. Happy, Sad, Anger, Fear, and Disgust are in between 1750-2000.

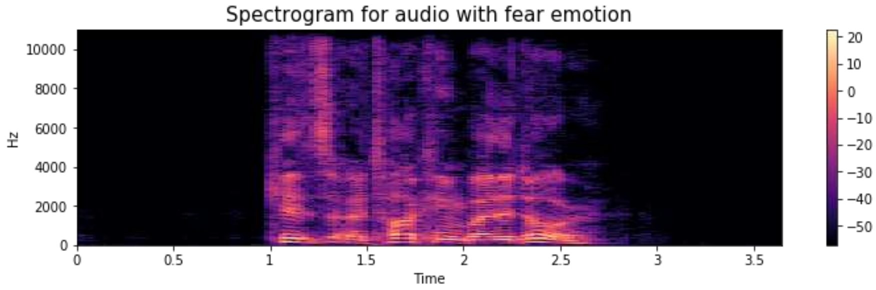
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



The waveplot for audio with fear emotion ranges from 1 to 2.5 at a given time.

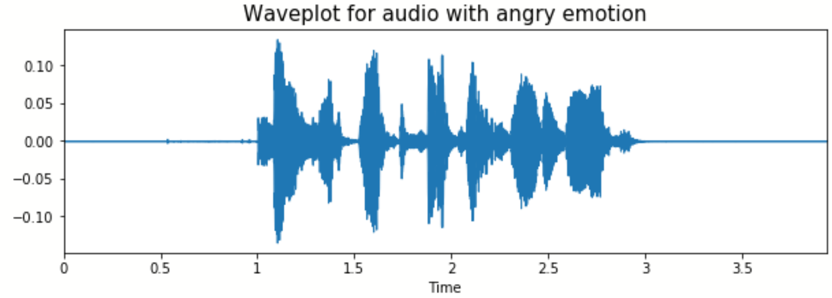


The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

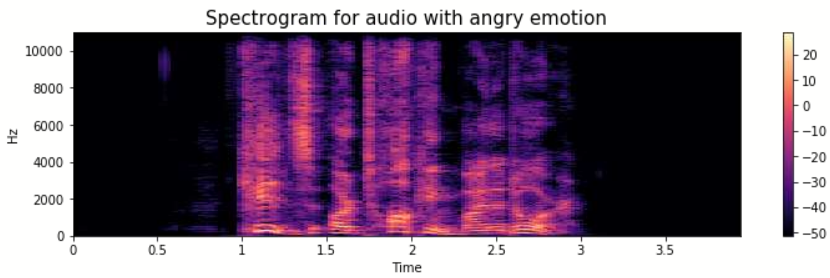


The waveplot for audio with angry emotion ranges from 1 to 3 at a given time.

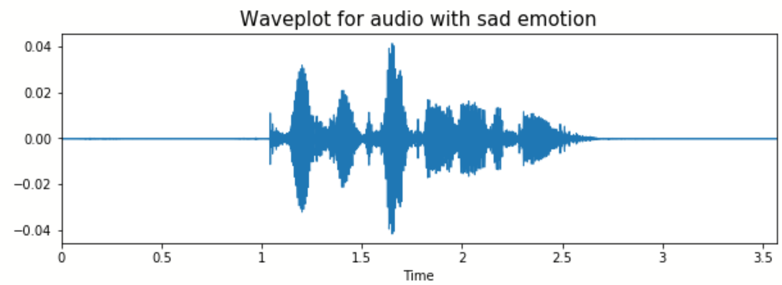
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



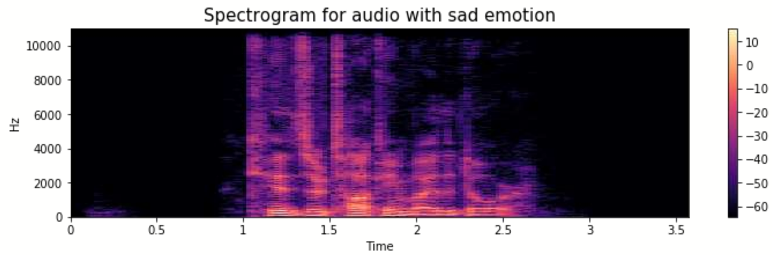
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



The waveplot for audio with sad emotion ranges from 1 to 3 at a given time.

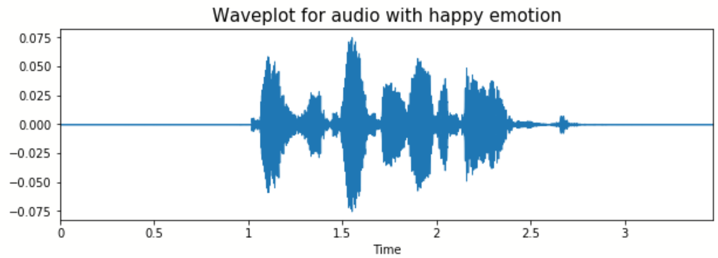


The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

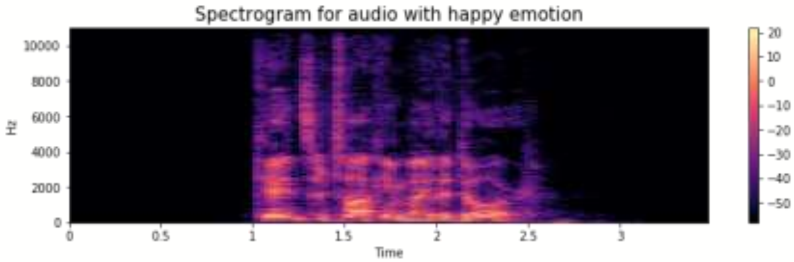


The waveplot for audio with happy emotion ranges from 1 to 3 at a given time.

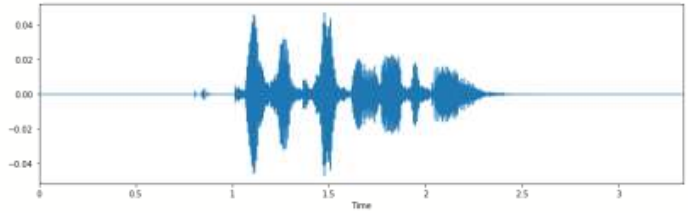
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

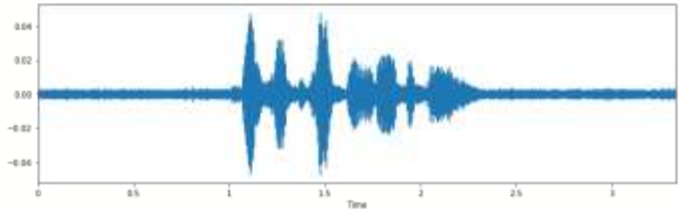


Let’s take any example and checking for techniques.

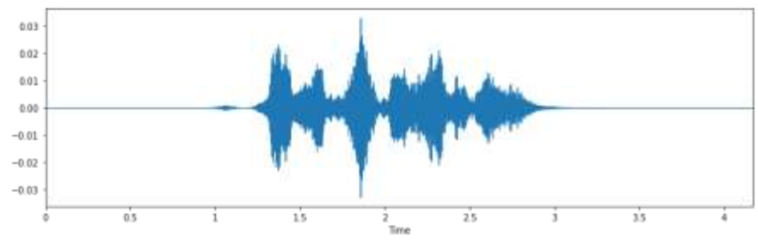


Noise injection is a very good augmentation technique because of which we can assure our training model is not over fitted.

x = noise(data)

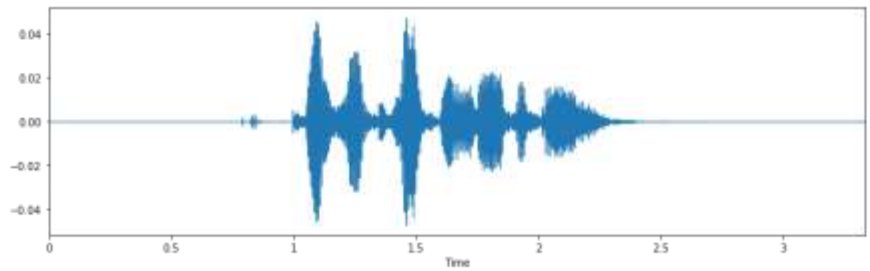


x = stretch(data)



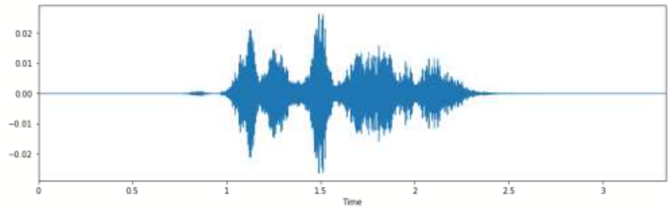
The above shows the stretch technique for the given data.

x = shift(data)



The above shows the shift technique for the given data.

x = pitch(data, sample_rate)



The above shows the pitch technique for the given data.

Predicted vs Actual

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

	Predicted Labels	Actual Labels
0	sad	disgust
1	disgust	disgust
2	angry	angry
3	happy	disgust
4	neutral	fear
5	sad	fear
6	angry	happy
7	happy	happy
8	neutral	sad
9	sad	sad

Confusion matrix:

A confusion matrix is used to measure the performance of a classifier in depth. Classification Models have multiple categorical outputs. Most error measures will calculate the total error in our model, but we cannot find individual instances of errors in our model. The model might misclassify some categories more than others, but we cannot see this using a standard accuracy measure.

Furthermore, suppose there is a significant class imbalance in the given data. In that case, i.e., a class has more instances of data than the other classes, a model might predict the majority class for all cases and have a high accuracy score; when it is not predicting the minority classes. This is where confusion matrices are useful.

A confusion matrix presents a table layout of the different outcomes of the prediction and results of a classification problem and helps visualize its outcomes. It plots a table of all the predicted and actual values of a classifier.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

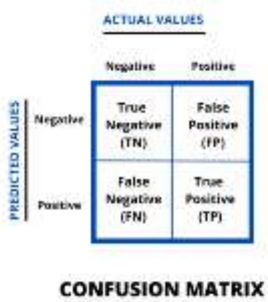
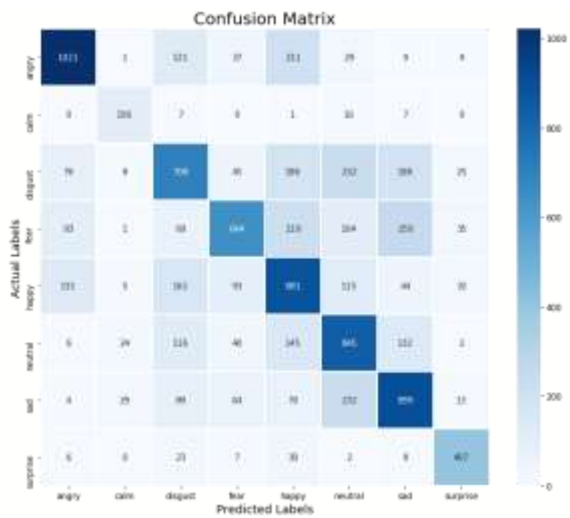


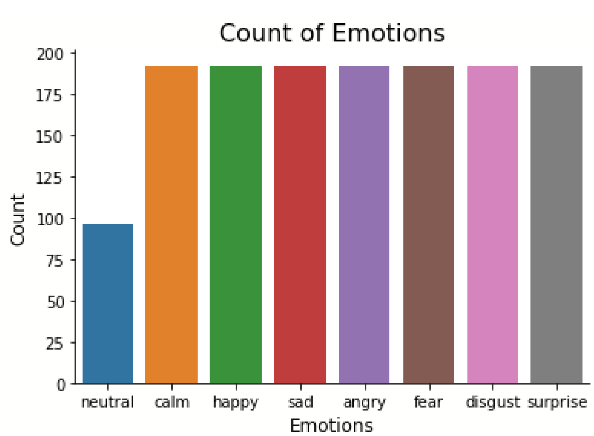
Fig 4.1 Confusion matrix

This is the confusion matrix



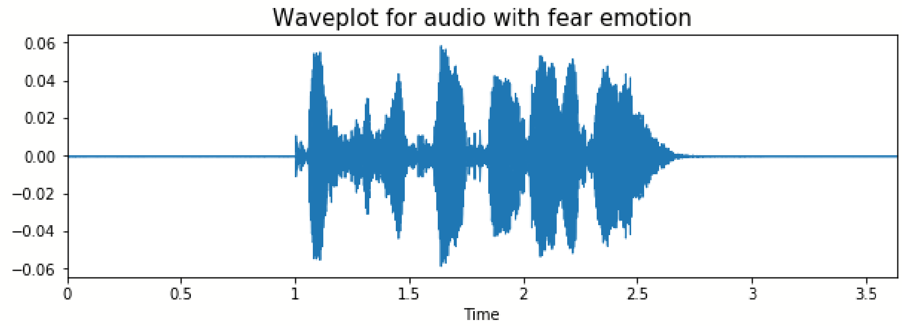
RAVDESS

The below bar graph shows the count of emotions. Neutral is between 75-100. Calm, Happy, Sad, Anger, Fear, Surprise, and Disgust are between 175-200.

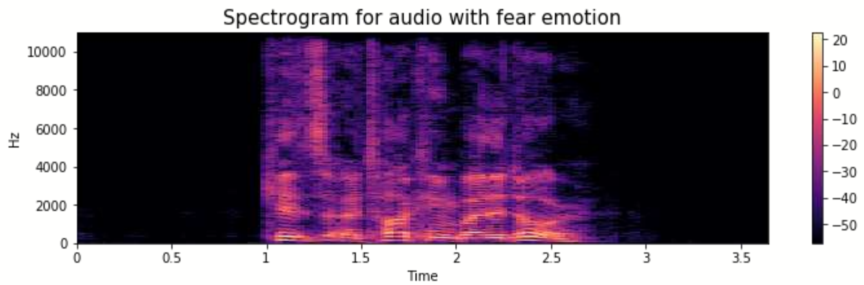


International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

The waveplot for audio with fear emotion ranges from 1 to 3 at a given time.

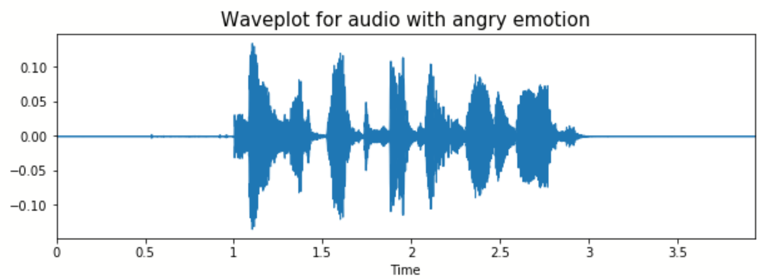


The spectrogram shows the visual representation of frequencies changing with respect to time for given audio

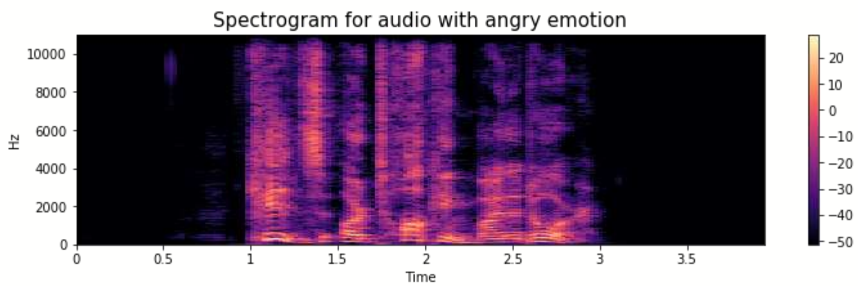


signals.

The waveplot for audio with angry emotion ranges from 1 to 3 at a given time.



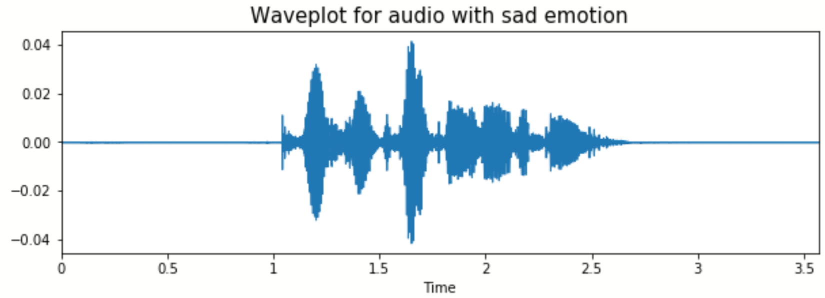
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio



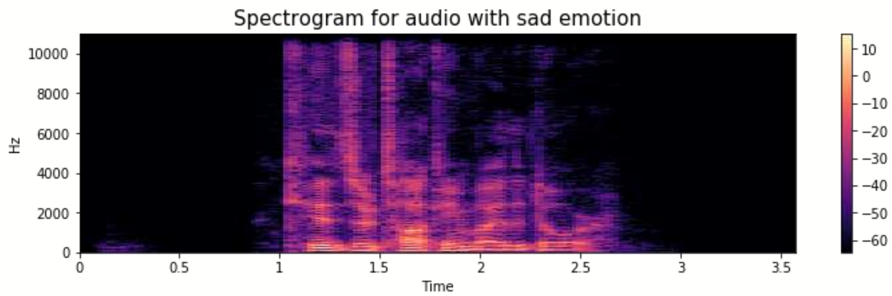
signals.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

The waveplot for audio with sad emotion ranges from 1 to 3 at a given time.

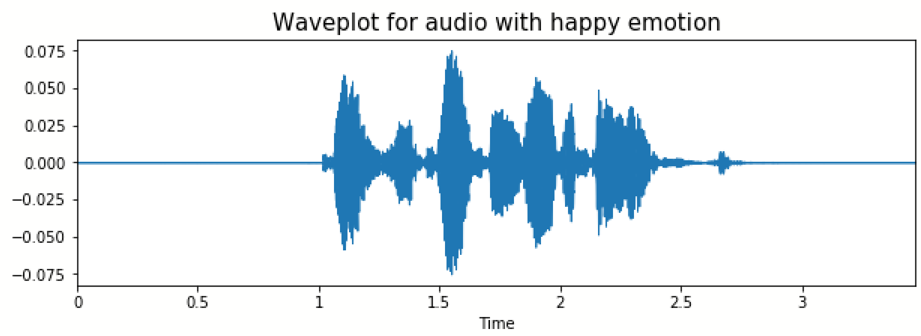


The spectrogram shows the visual representation of frequencies changing with respect to time for given audio

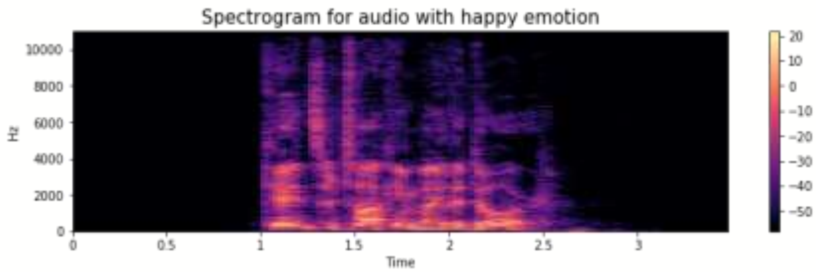


signals.

The waveplot for audio with happy emotion ranges from 1 to 3 at a given time.



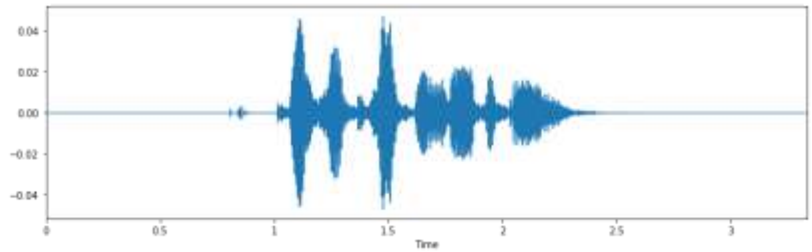
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio



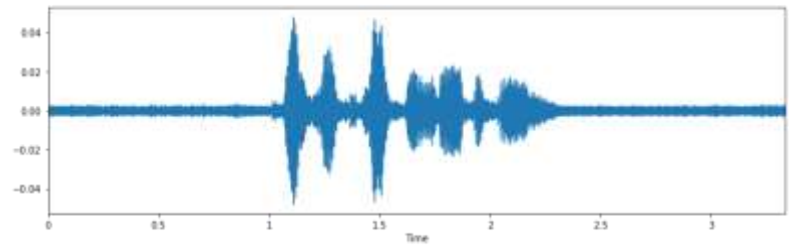
signals.

Let’s take any example and checking for techniques.

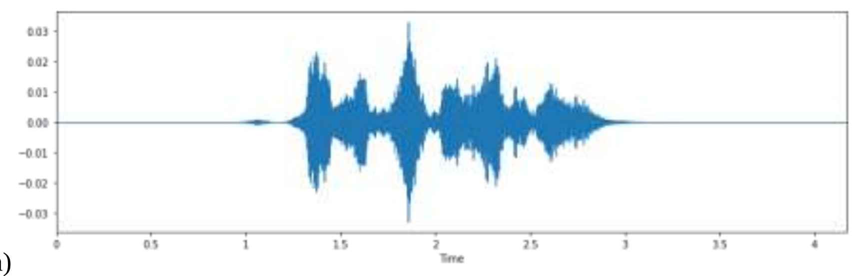
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



x = noise(data)



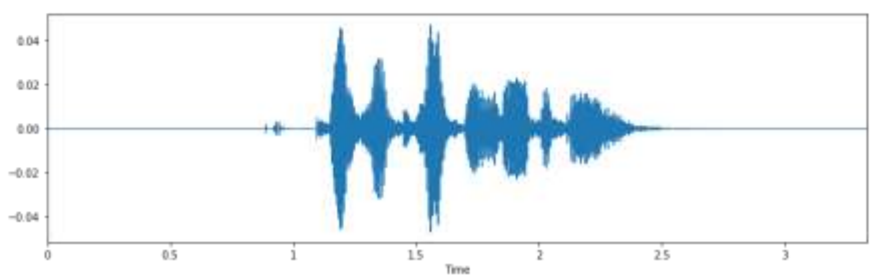
Noise injection is a very good augmentation technique because of which we can assure our training model is not overfitted.



x = stretch(data)

The above shows the stretch technique for the given data.

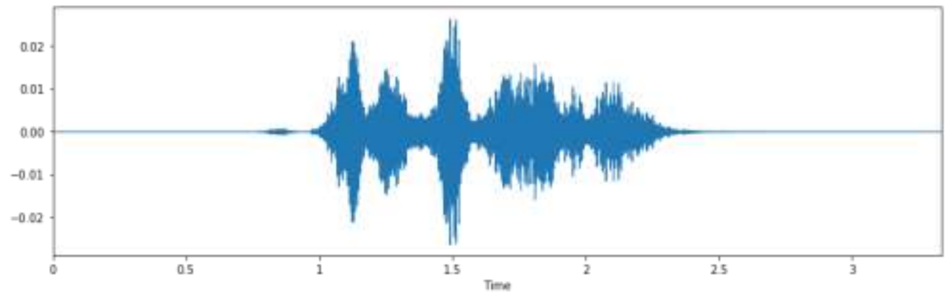
x = shift(data)



The above shows the shift technique for the given data.

x = pitch(data, sample_rate)

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



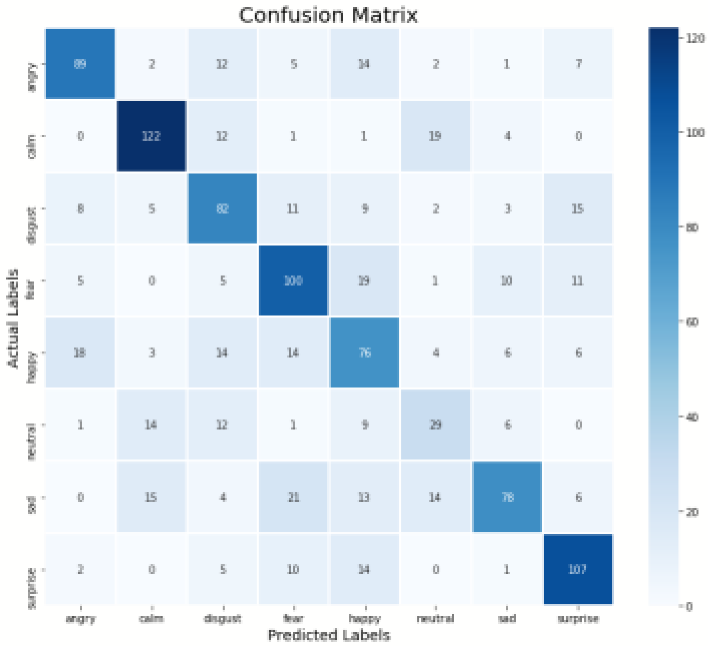
The above shows the pitch technique for the given data.

Predicted vs Actual

	Predicted Labels	Actual Labels
0	fear	fear
1	fear	angry
2	fear	fear
3	calm	calm
4	angry	angry
5	surprise	surprise
6	sad	fear
7	happy	happy
8	fear	fear
9	sad	sad

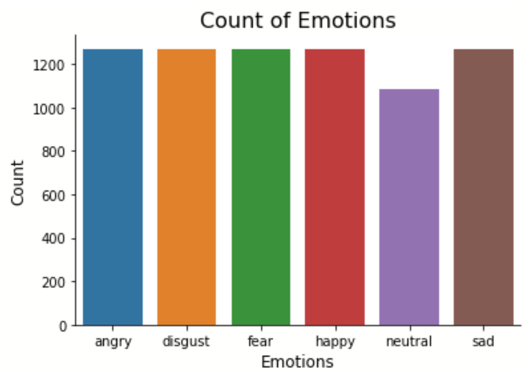
This is the confusion matrix

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

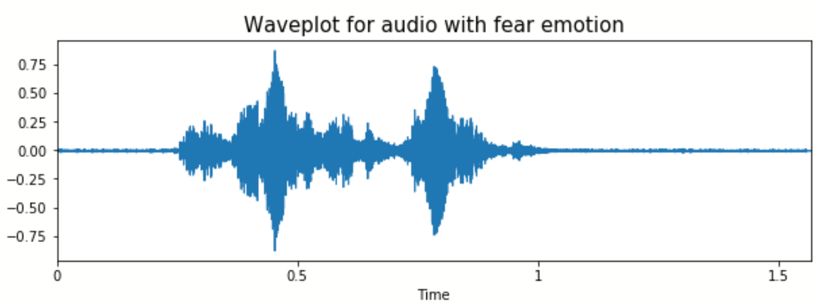


CREMA

The below bar graph shows the count of emotions. Angry, Disgust, Fear, Happy, and Sad are above 1200. Neutral is between 1000-1200.

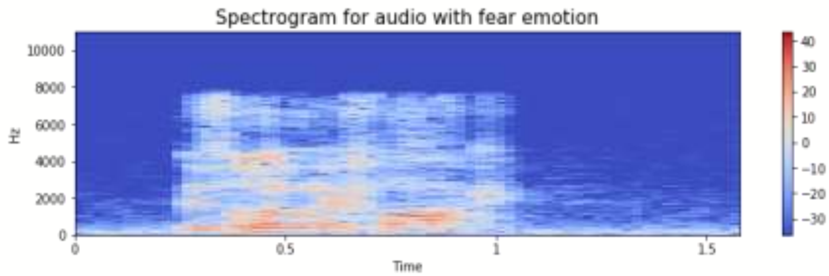


The waveplot for audio with fear emotion ranges from 0 to 1 at a given time.

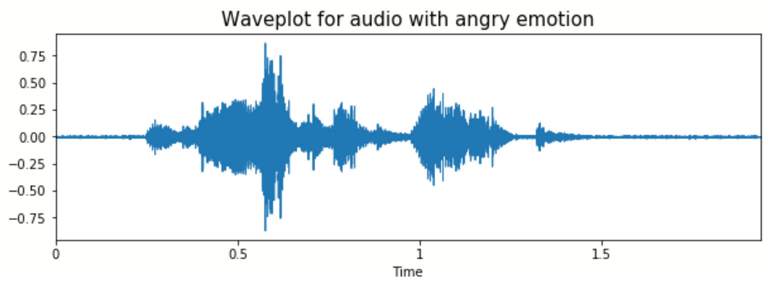


International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

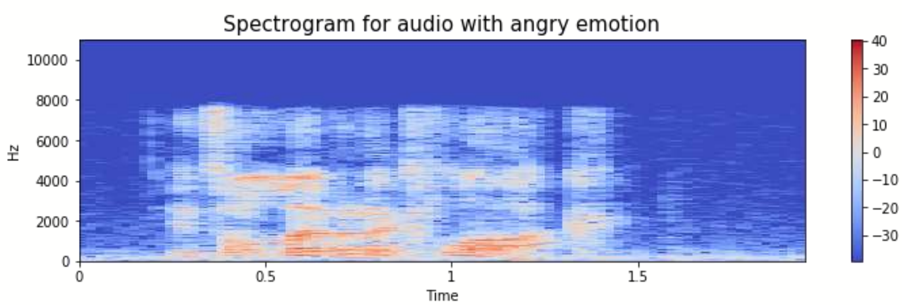
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



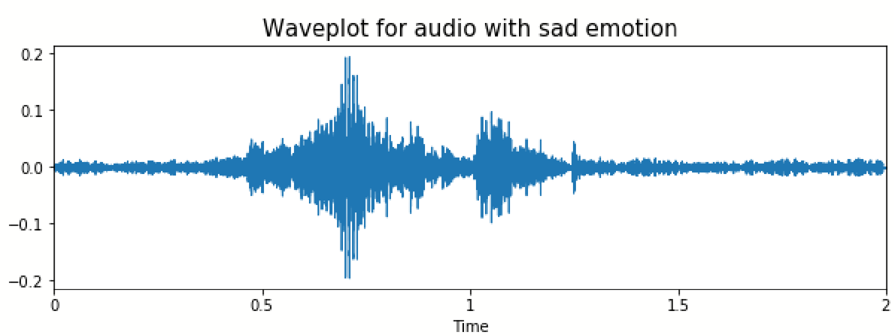
The waveplot for audio with angry emotion ranges from 0 to 1.5 at a given time.



The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

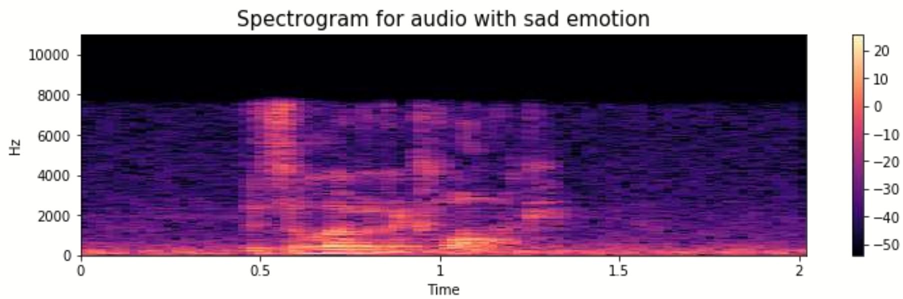


The waveplot for audio with sad emotion ranges from 0 to 1.5 at a given time.



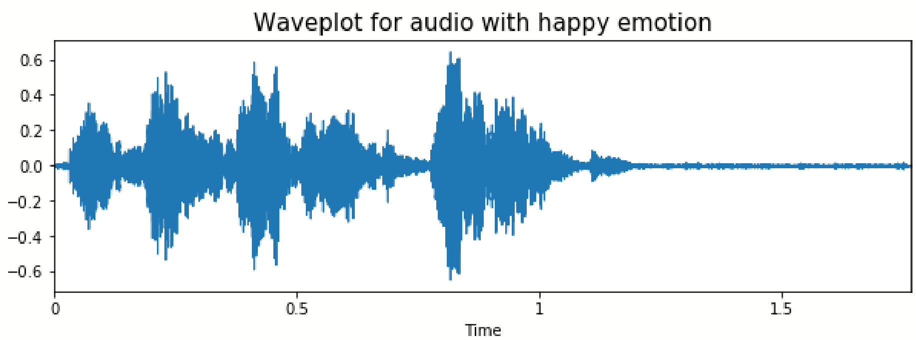
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

The spectrogram shows the visual representation of frequencies changing with respect to time for given audio

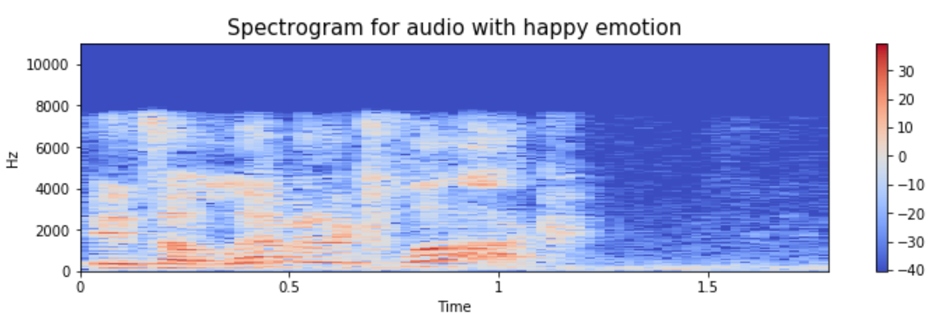


signals.

The waveplot for audio with happy emotion ranges from 0 to 1 at a given time.

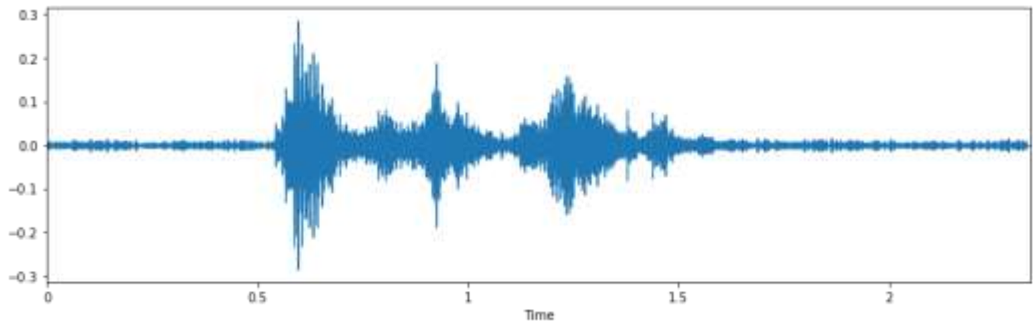


The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



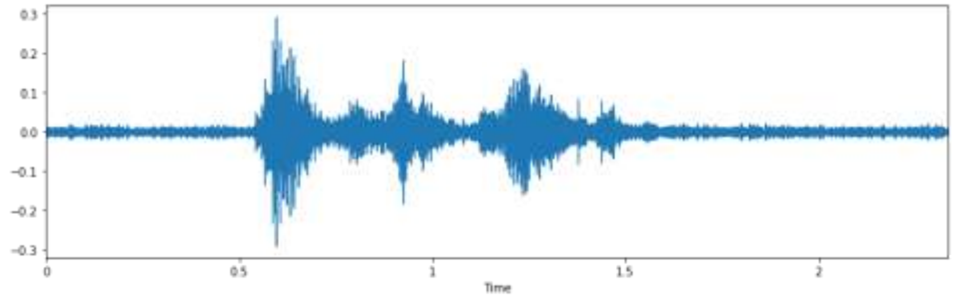
Let’s take any example and checking for techniques.

International Journal for Innovative
Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

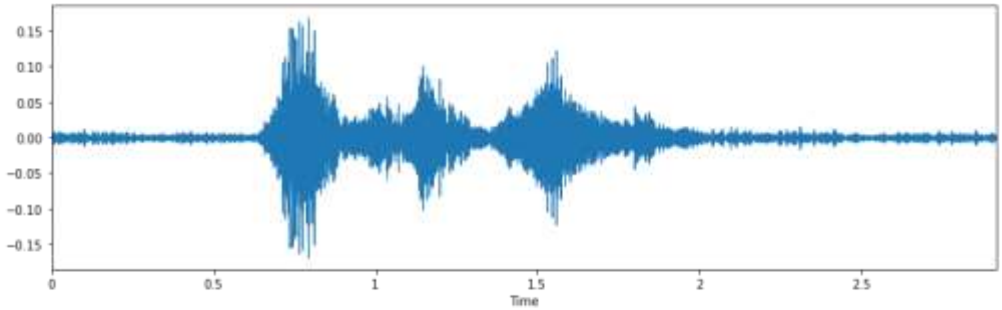


Noise injection is a very good augmentation technique because of which we can assure our training model is not overfitted.

x = noise(data)

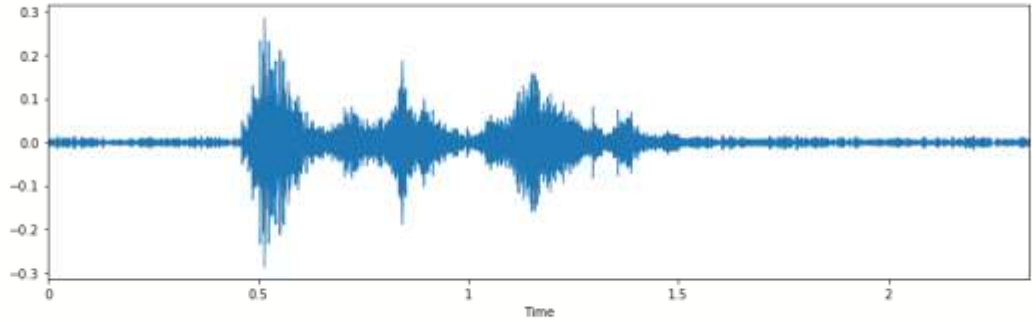


x = stretch(data)



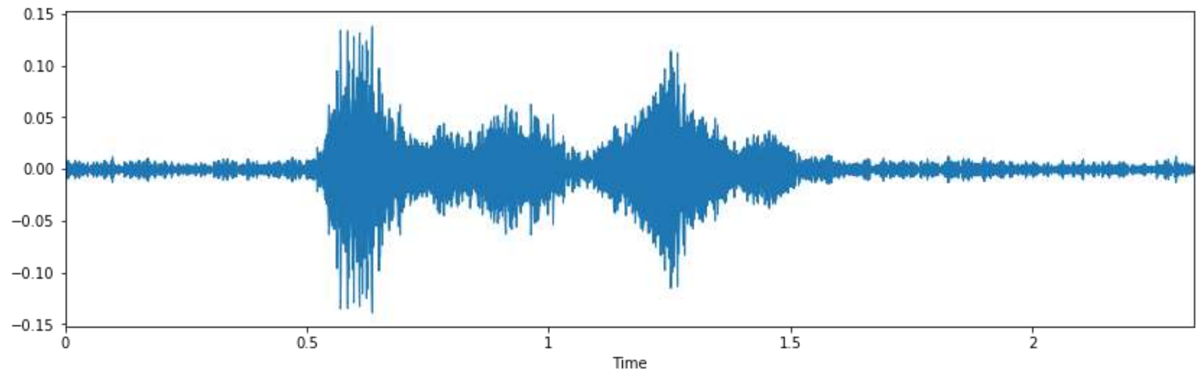
The above shows the stretch technique for the given data.

x = shift(data)



The above shows the shift technique for the given data.

x = pitch(data, sample_rate)



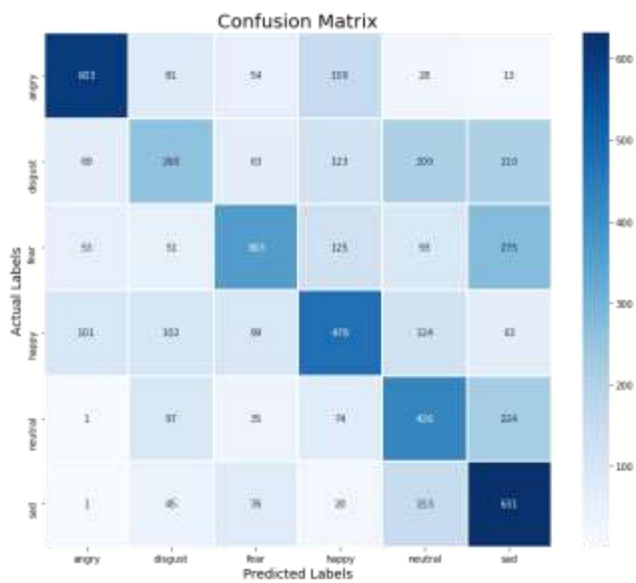
The above shows the pitch technique for the given data.

Predicted Vs Actual

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

	Predicted Labels	Actual Labels
0	sad	fear
1	happy	happy
2	sad	sad
3	sad	neutral
4	disgust	disgust
5	neutral	neutral
6	angry	angry
7	sad	sad
8	happy	neutral
9	sad	sad

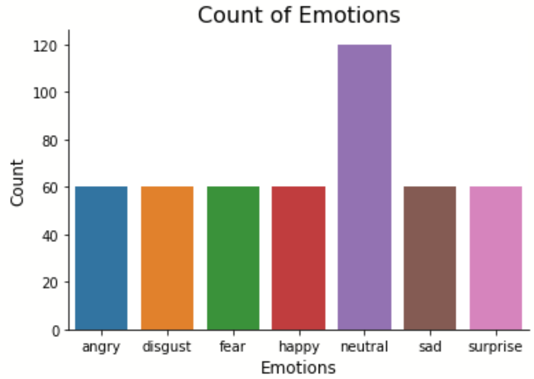
This is the confusion matrix



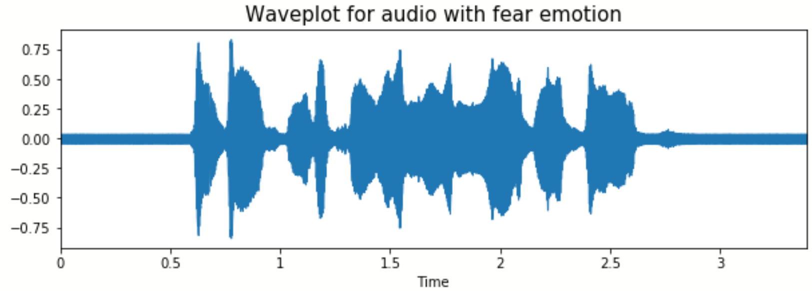
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

SAVEE

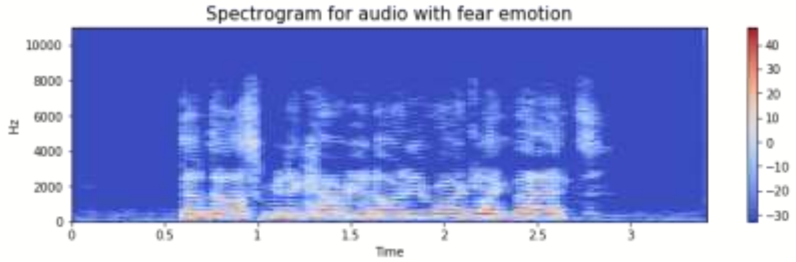
The below bar graph shows the count of emotions. Angry, Disgust, Fear, Happy, Sad, and Surprise are in 60. Neutral is 120.



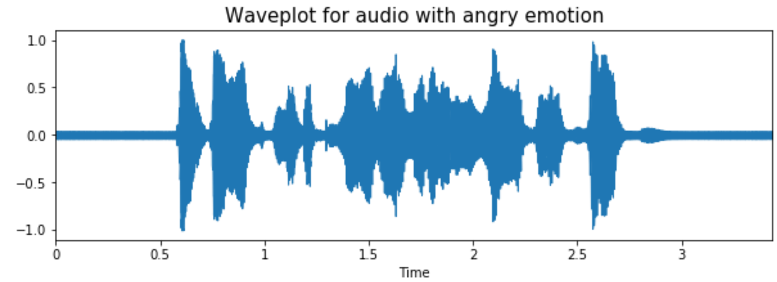
The waveplot for audio with fear emotion ranges from 1 to 3 at a given time.



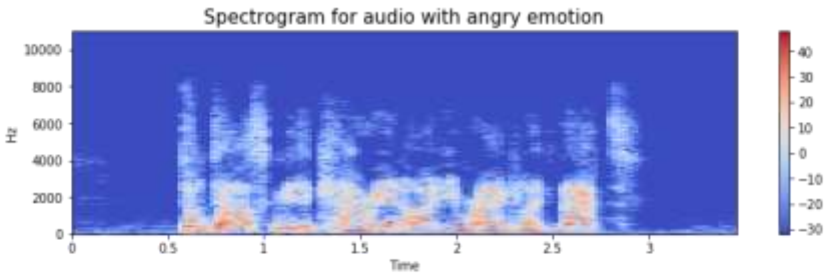
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



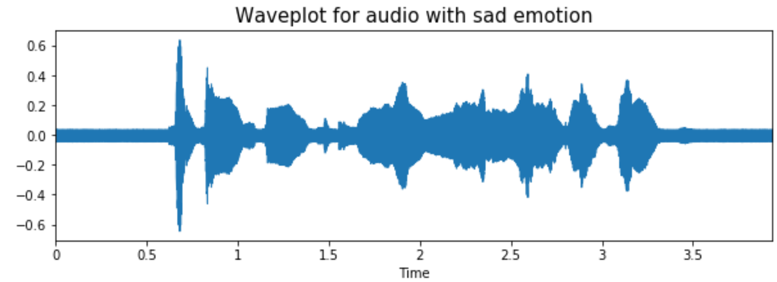
The waveplot for audio with angry emotion ranges from 1 to 3 at a given time.



The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

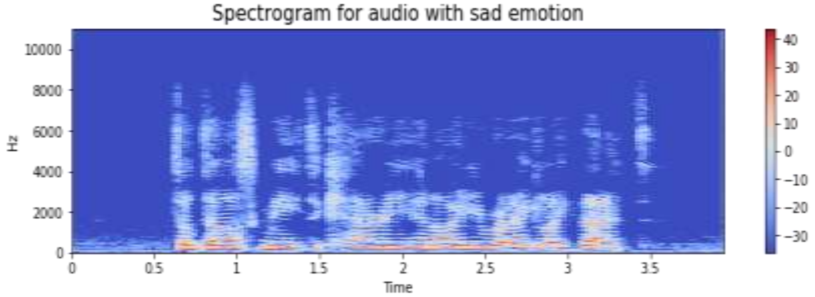


The waveplot for audio with sad emotion ranges from 0.5 to 3.5 at a given time.

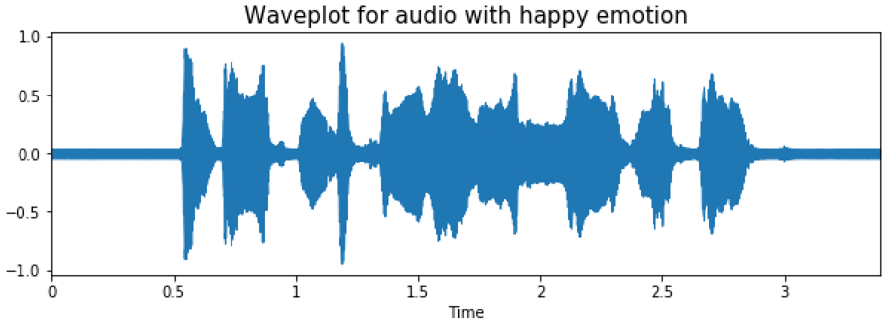


The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

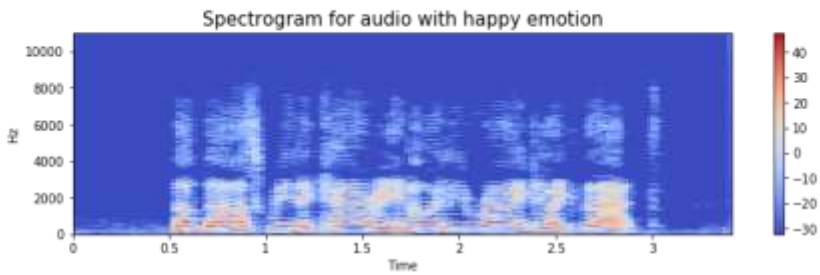
The waveplot for audio with happy emotion ranges from 0.5 to 3 at a given time.



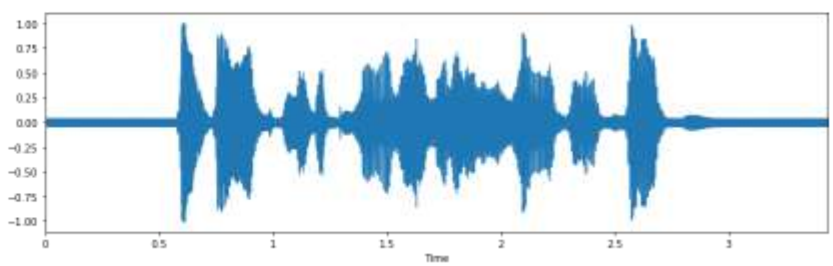
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

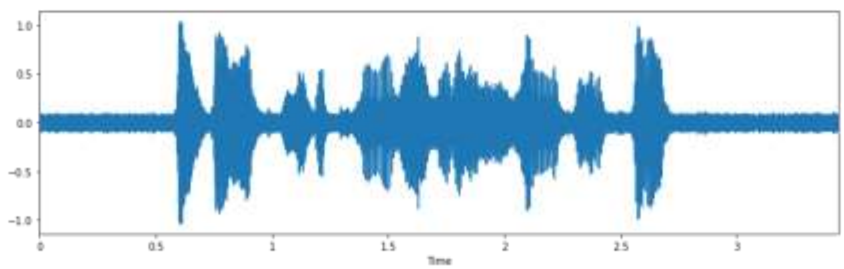


Let's take any example and checking for techniques.



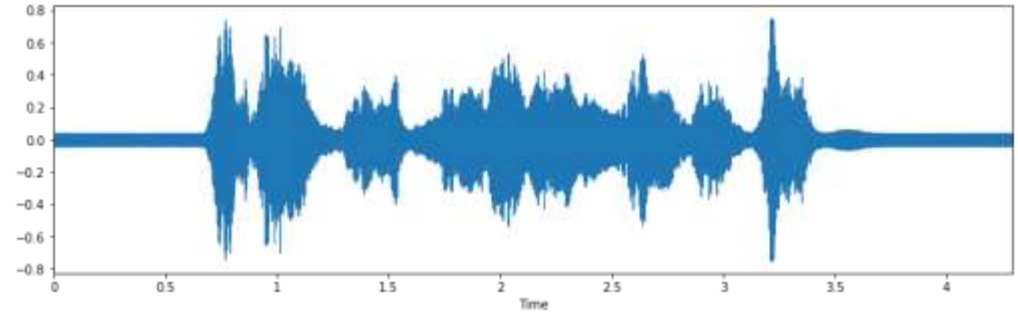
Noise injection is a very good augmentation technique because of which we can assure our training model is not overfitted.

x = noise(data)



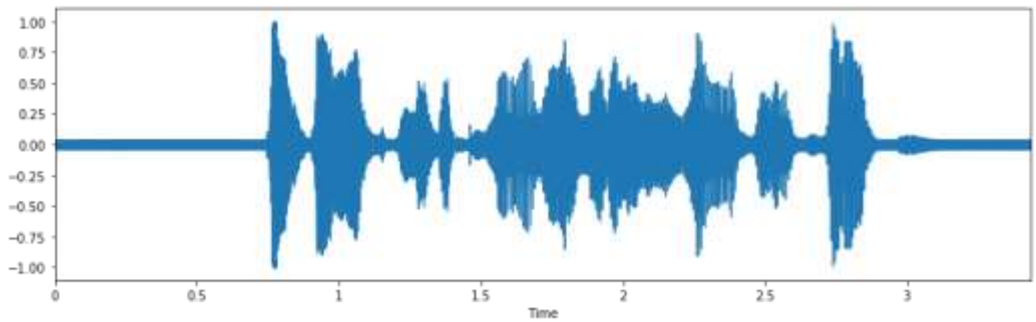
x = stretch(data)

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



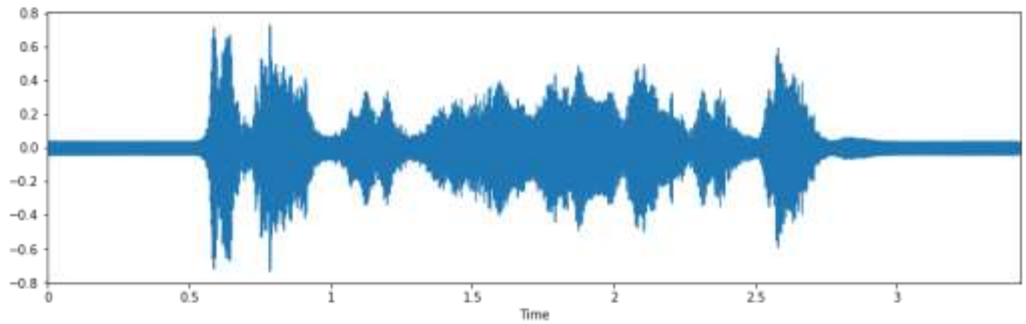
The above shows the stretch technique for the given data.

x = shift(data)



The above shows the shift technique for the given data.

x = pitch(data, sample_rate)



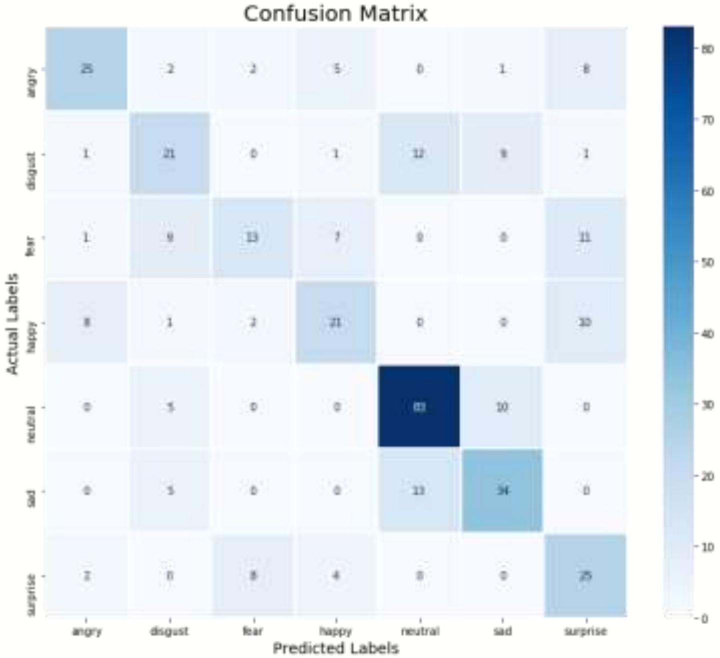
The above shows the pitch technique for the given data.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Predicted vs Actual

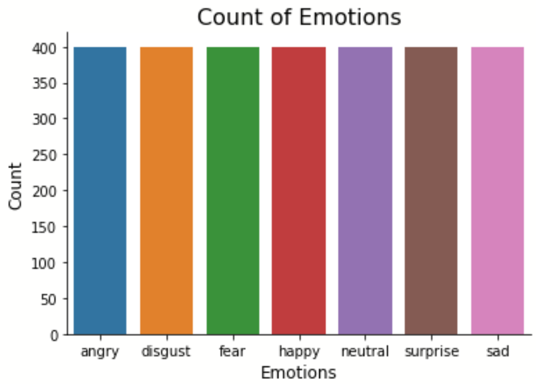
	Predicted Labels	Actual Labels
0	sad	angry
1	surprise	surprise
2	surprise	happy
3	happy	surprise
4	happy	happy
5	angry	angry
6	neutral	neutral
7	sad	sad
8	disgust	fear
9	disgust	disgust

This is the confusion matrix

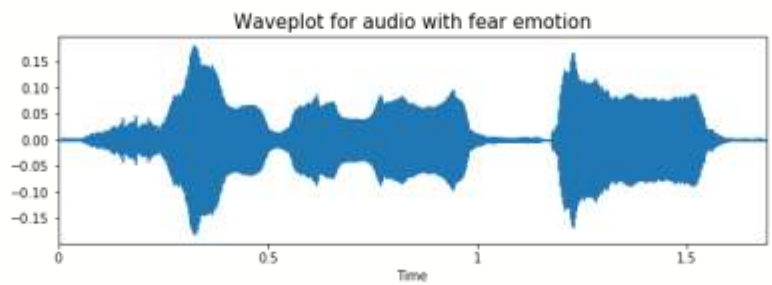


TESS

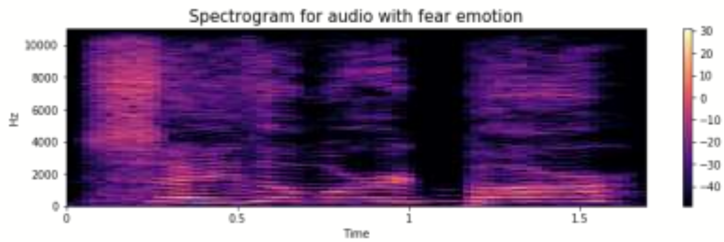
The below bar graph shows the count of emotions. Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral is 400.



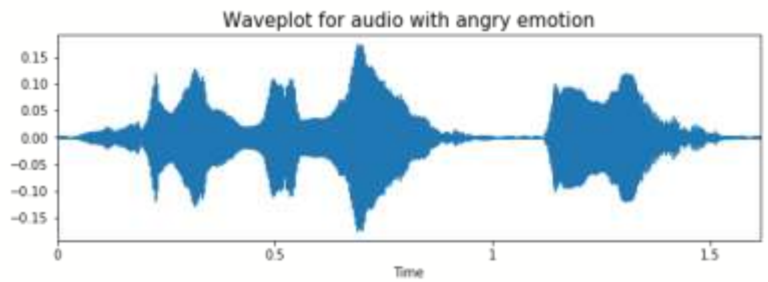
The waveplot for audio with fear emotion ranges from 0 to 2 at a given time.



The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

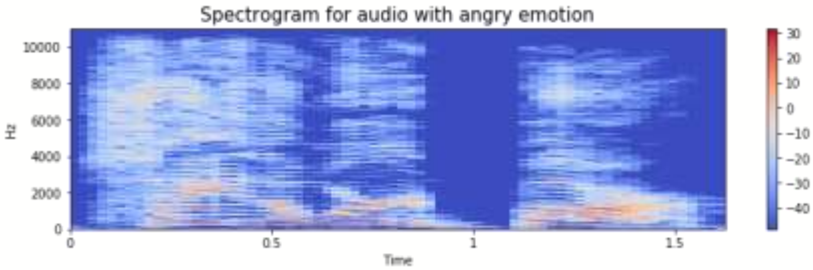


The waveplot for audio with angry emotion ranges from 0 to 1.5 at a given time.

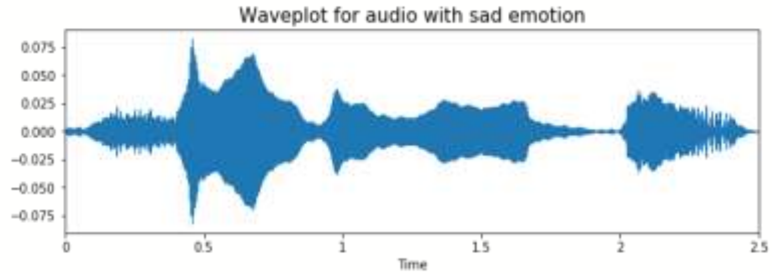


International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

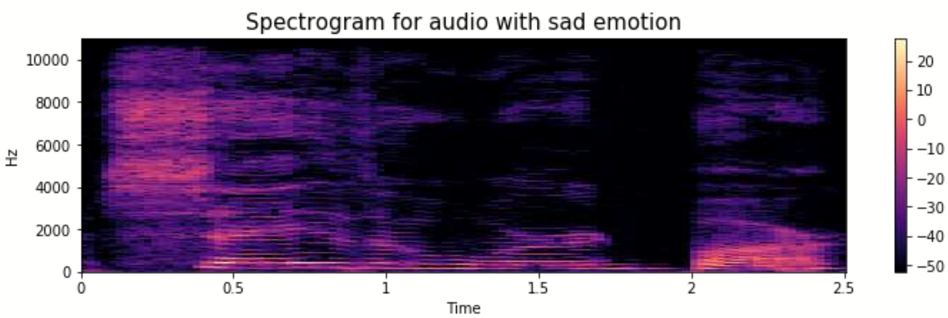


The waveplot for audio with sad emotion ranges from 1 to 2.5 at a given time.



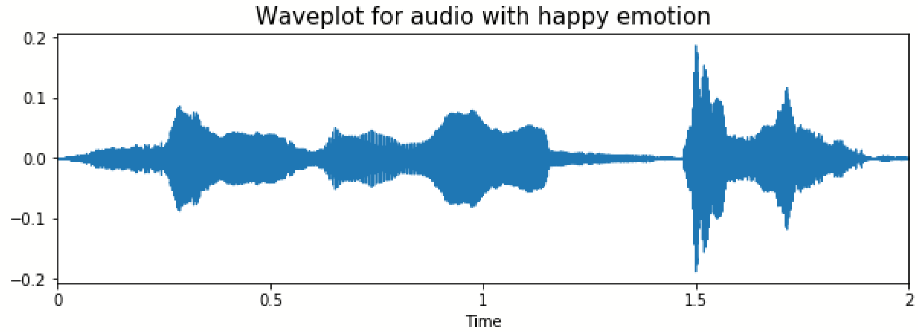
\

The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

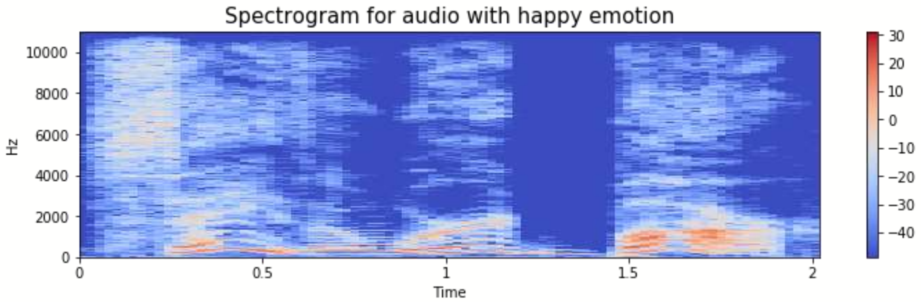


The waveplot for audio with happy emotion ranges from 0 to 2 at a given time.

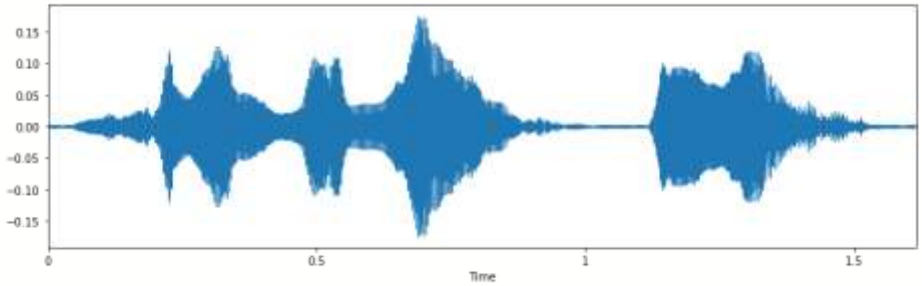
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



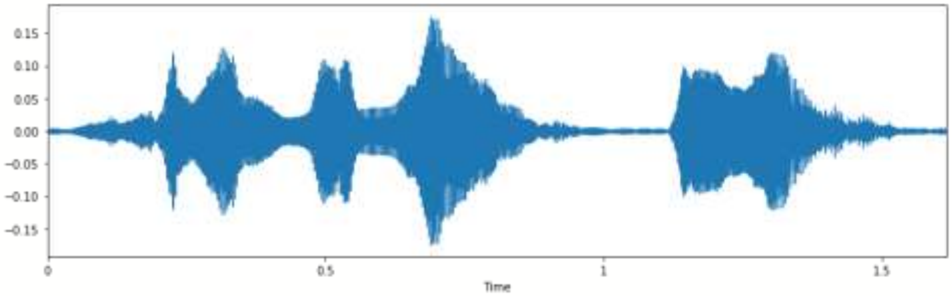
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals



Let’s take any example and checking for techniques.



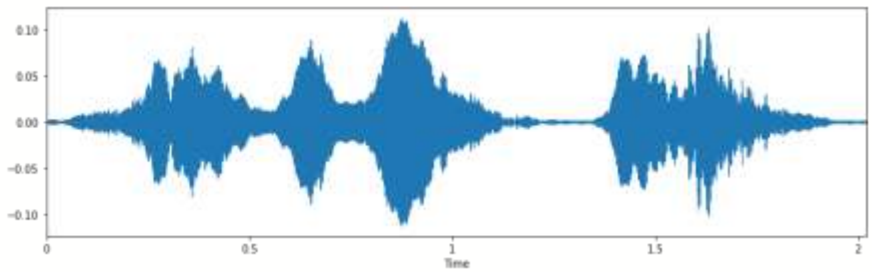
x = noise(data)



Noise injection is a very good augmentation technique because of which we can assure our training model is not overfitted.

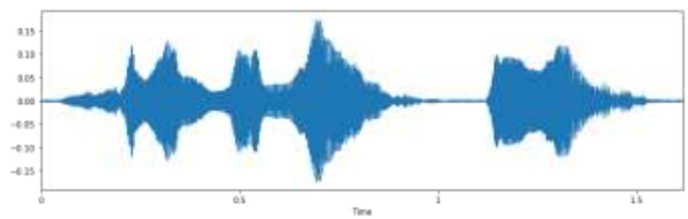
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

x = stretch(data)

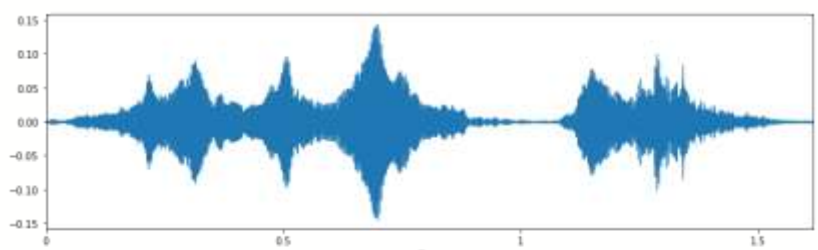


The above shows the stretch technique for the given data.

x = shift(data)



The above shows the shift technique for the given data.



x = pitch(data, sample_rate)

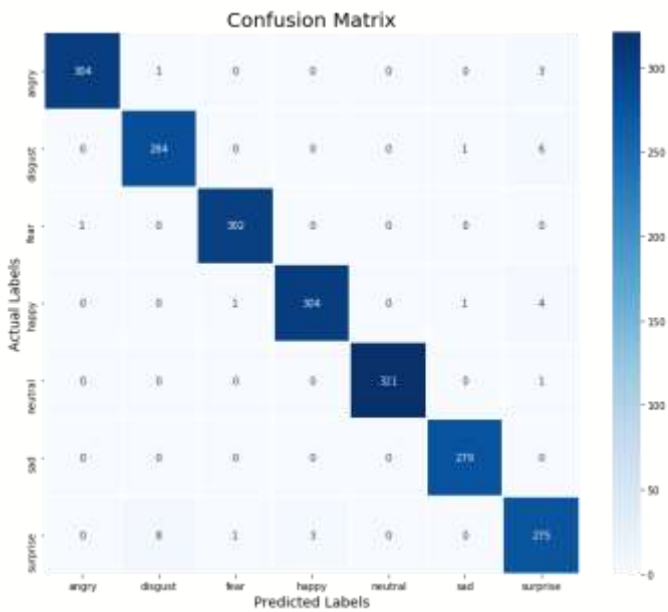
The above shows the pitch technique for the given data.

Predicted vs Actual

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

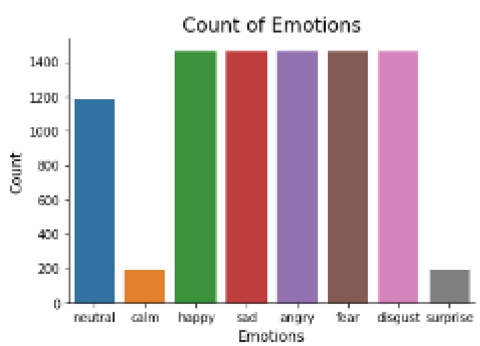
	Predicted Labels	Actual Labels
0	neutral	neutral
1	surprise	surprise
2	neutral	neutral
3	happy	happy
4	surprise	surprise
5	fear	fear
6	fear	fear
7	sad	sad
8	disgust	disgust
9	fear	fear

This is the confusion matrix

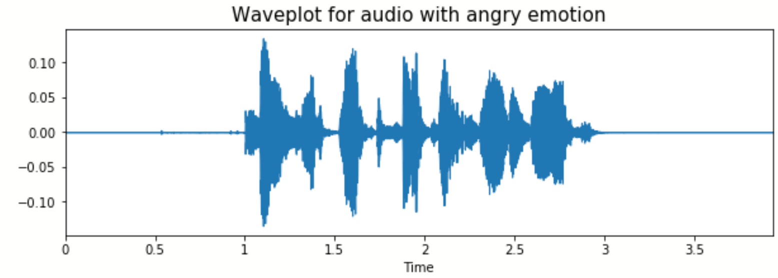


RAVDESS+CREMA

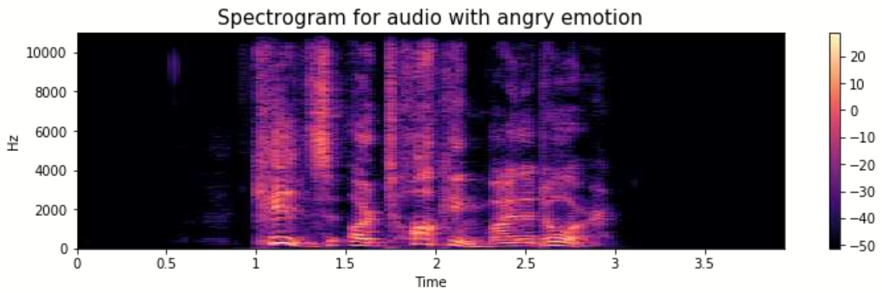
The below bar graph shows the count of emotions. Netural is between 0-1200, whereas calm and surprise is between 0-200. Happy, Sad, Anger, Fear, and Disgust are in between 1400



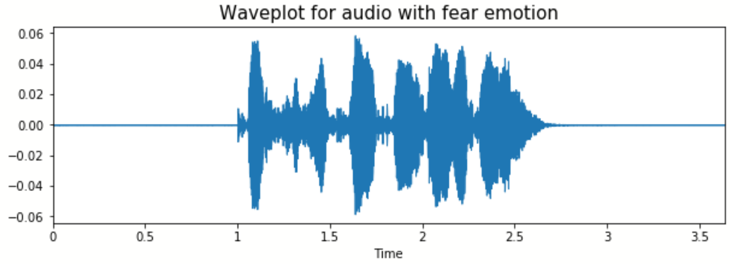
The waveplot for audio with fear emotion ranges from 0 to 3 at a given time.



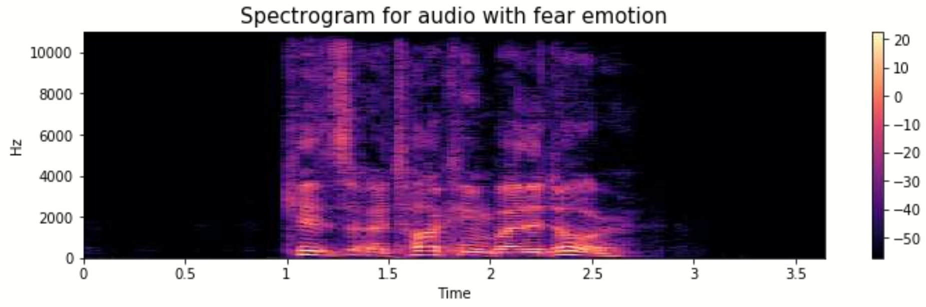
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



The waveplot for audio with angry emotion ranges from 1 to 3 at a given time.

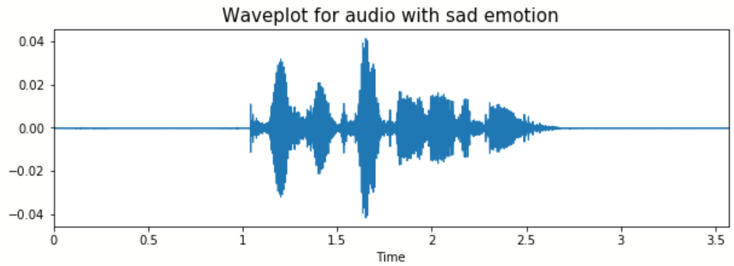


The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

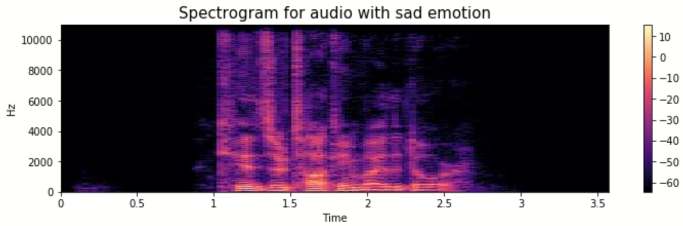


The waveplot for audio with sad emotion ranges from 1 to 2.5 at a given time.

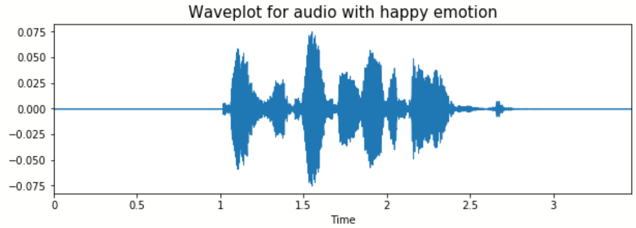
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



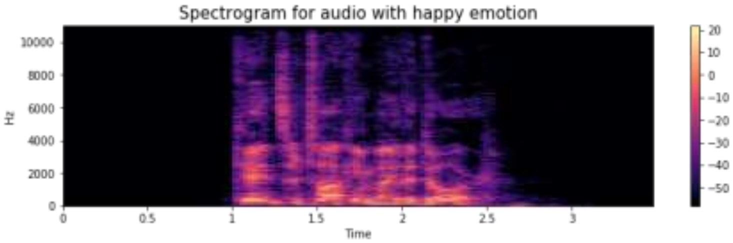
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



The waveplot for audio with happy emotion ranges from 1 to 3 at a given time.

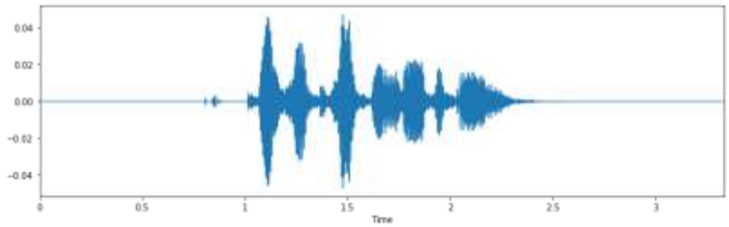


The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



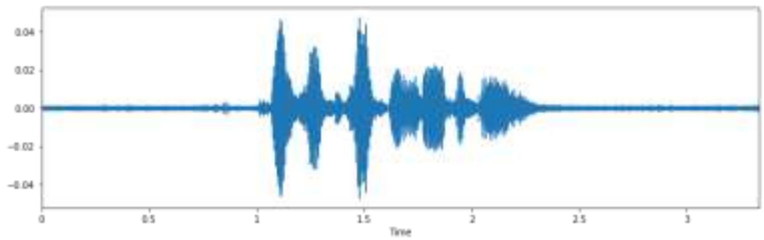
Let's take any example and checking for techniques.

International Journal for Innovative
Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

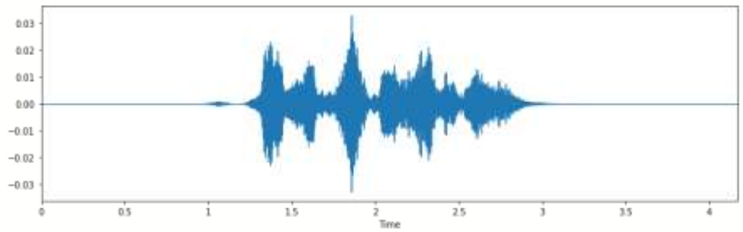


x = noise(data)

Noise injection is a very good augmentation technique because of which we can assure our training model is not overfitted.

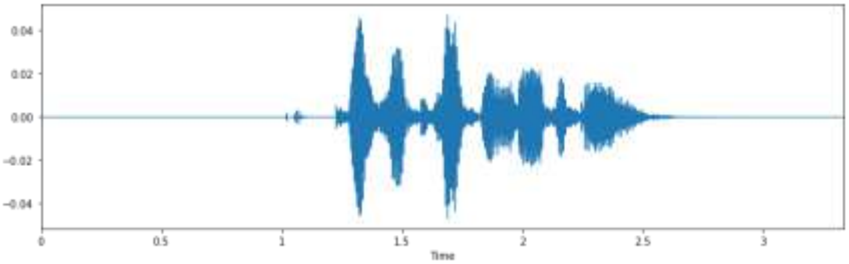


x = stretch(data)



The above shows the stretch technique for the given data.

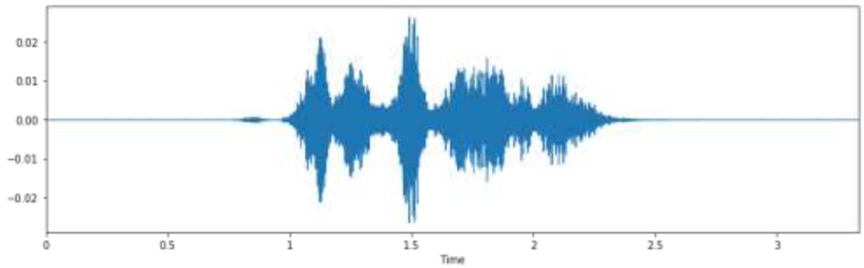
x = shift(data)



The above shows the shift technique for the given data.

x = pitch(data, sample_rate)

International Journal for Innovative
Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

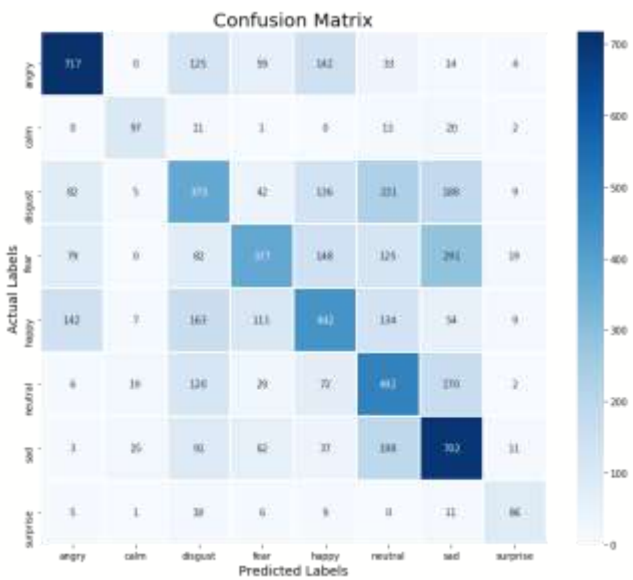


The above shows the pitch technique for the given data.

Predicted vs Actual

	Predicted Labels	Actual Labels
0	happy	neutral
1	disgust	fear
2	angry	angry
3	disgust	neutral
4	neutral	angry
5	sad	sad
6	sad	neutral
7	calm	calm
8	sad	sad
9	sad	sad

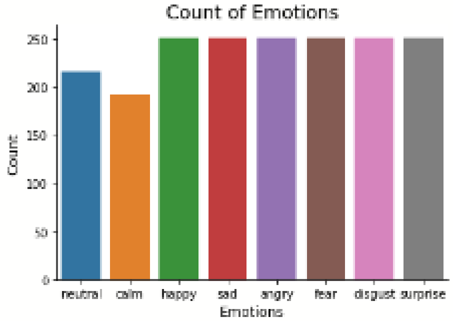
This is the confusion matrix



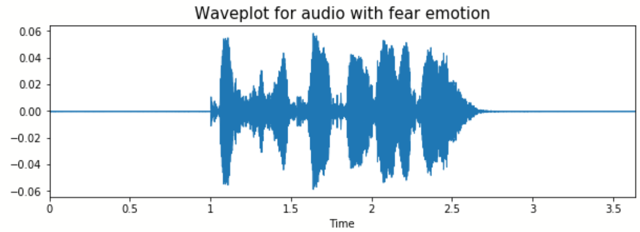
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

RAVDESS+SAVEE

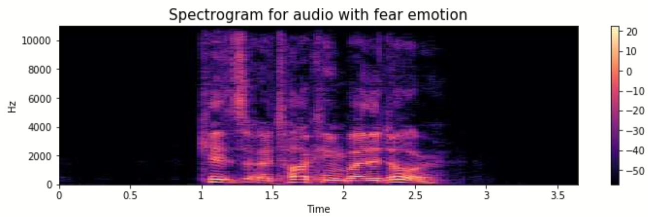
The below bar graph shows the count of emotions. Neutral is between 200-250, whereas calm is between 150-200. Happy, Sad, Anger, Fear, and Disgust and Surprise is 250



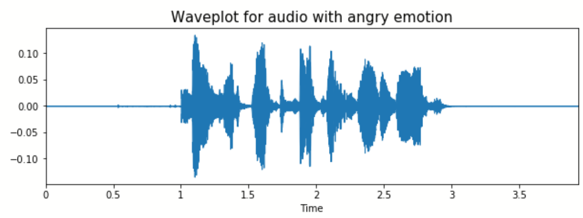
The waveplot for audio with fear emotion ranges from 1 to 2.5 at a given time.



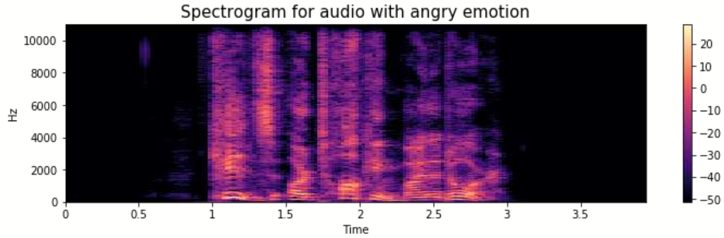
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



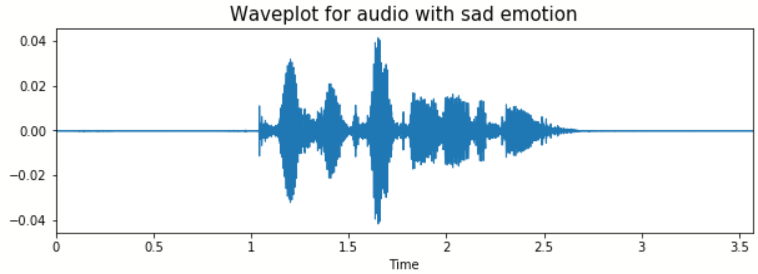
The waveplot for audio with angry emotion ranges from 1 to 3 at a given time.



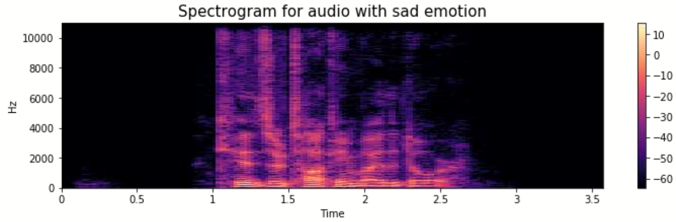
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



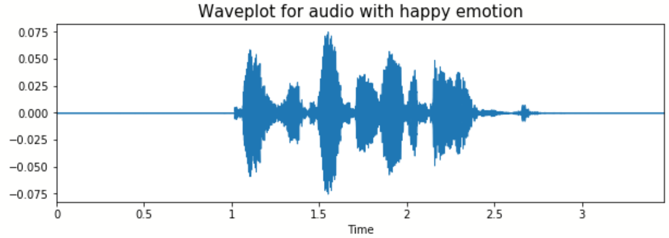
The waveplot for audio with sad emotion ranges from 1 to 2.5 at a given time.



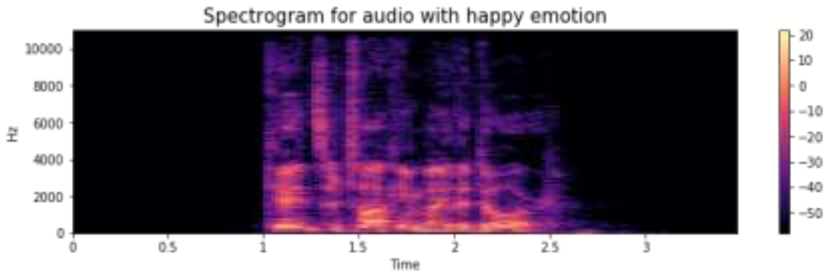
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



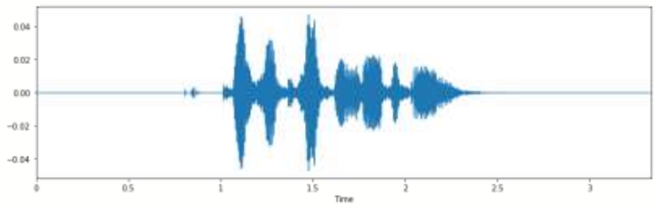
The waveplot for audio with happy emotion ranges from 1 to 2.5 at a given time.



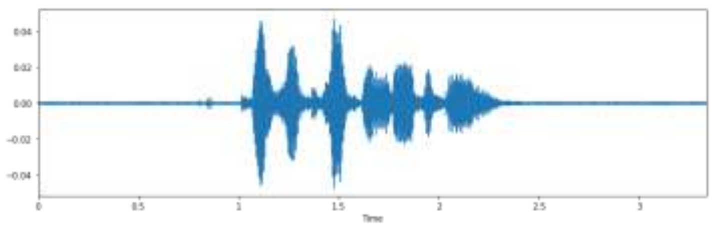
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



Let’s take any example and checking for techniques.

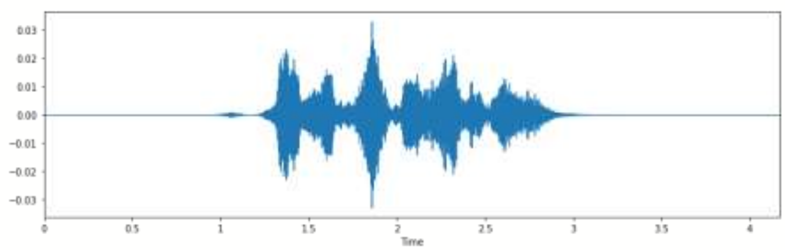


x = noise(data)



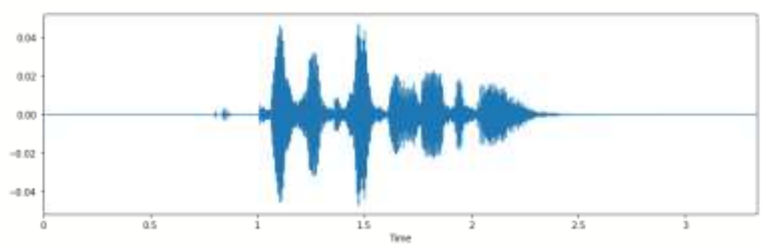
Noise injection is a very good augmentation technique because of which we can assure our training model is not overfitted.

x = stretch(data)



The above shows the stretch technique for the given data.

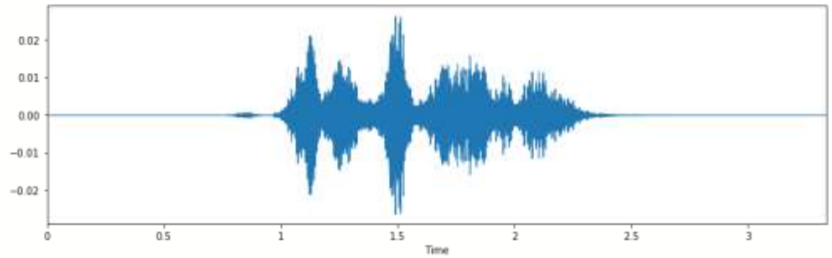
x = shift(data)



International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

The above shows the shift technique for the given data.

x = pitch(data, sample_rate)

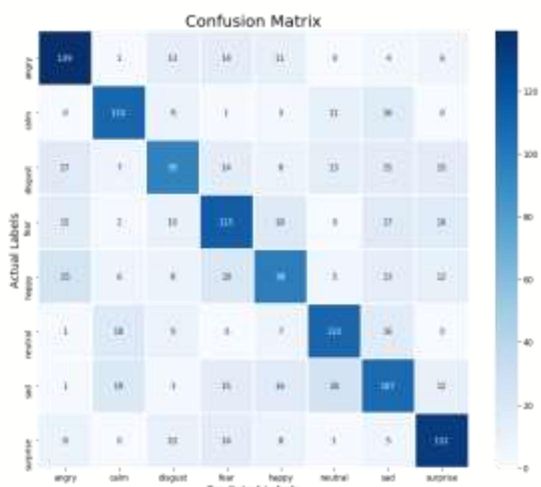


The above shows the pitch technique for the given data

Predicted vs Actual

	Predicted Labels	Actual Labels
0	happy	angry
1	fear	angry
2	sad	sad
3	disgust	disgust
4	calm	calm
5	fear	fear
6	surprise	happy
7	sad	sad
8	fear	fear
9	fear	happy

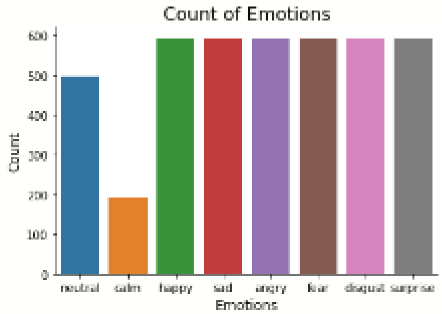
Confusion Matrix



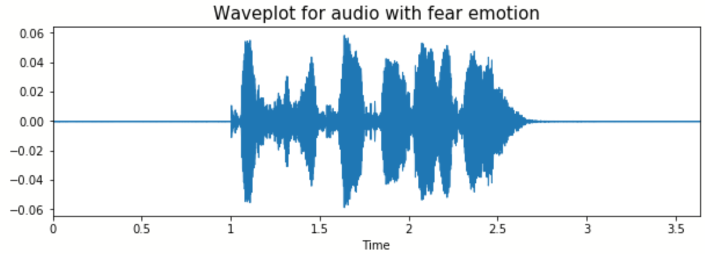
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

RAVDESS+TESS

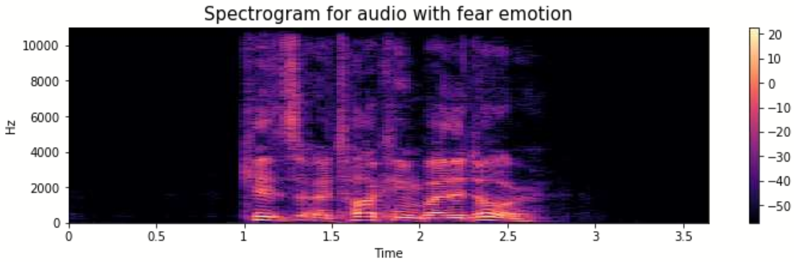
The below bar graph shows the count of emotions. Neutral is between 500, whereas calm is between 100-200. Happy, Sad, Anger, Fear, Disgust and Surprise are in between 600.



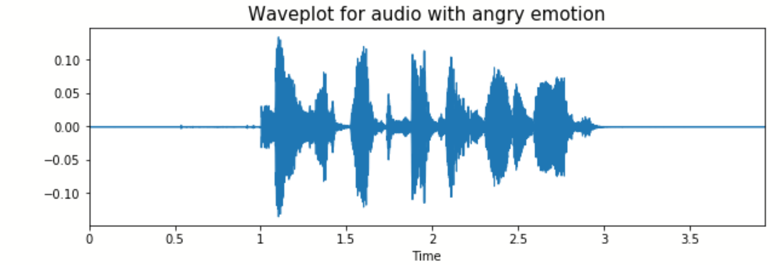
The waveplot for audio with fear emotion ranges from 1 to 2.5 at a given time.



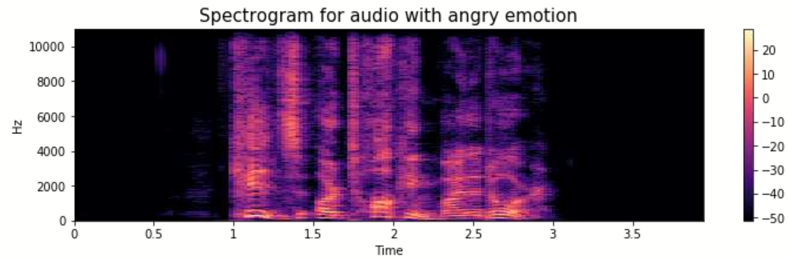
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



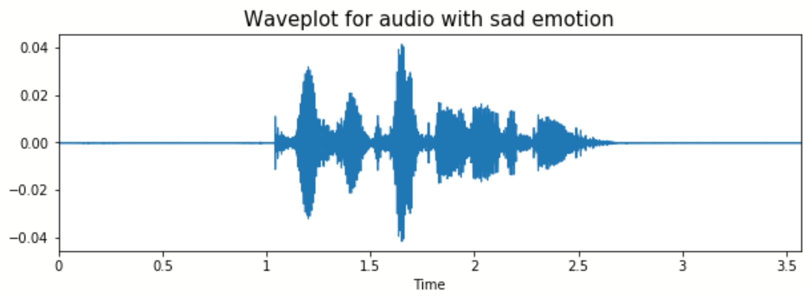
The waveplot for audio with angry emotion ranges from 1 to 3 at a given time.



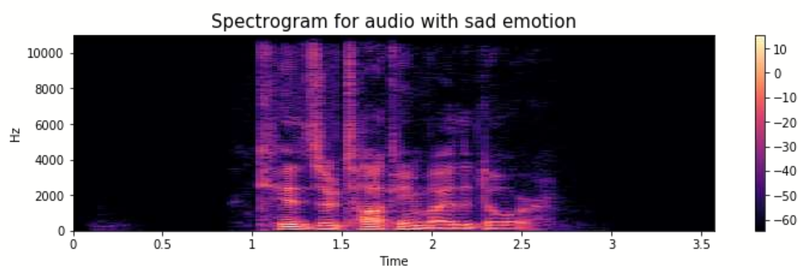
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



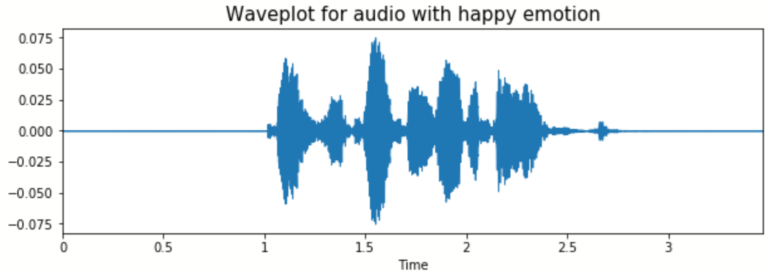
The waveplot for audio with sad emotion ranges from 1 to 2.5 at a given time.



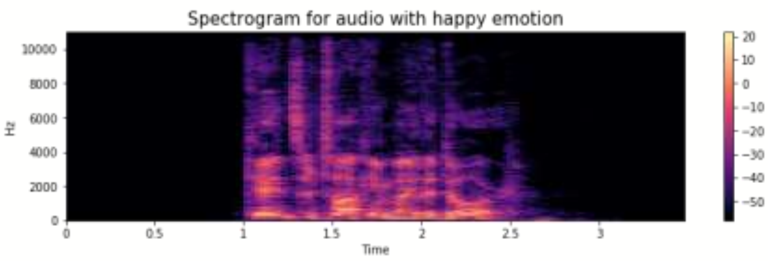
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



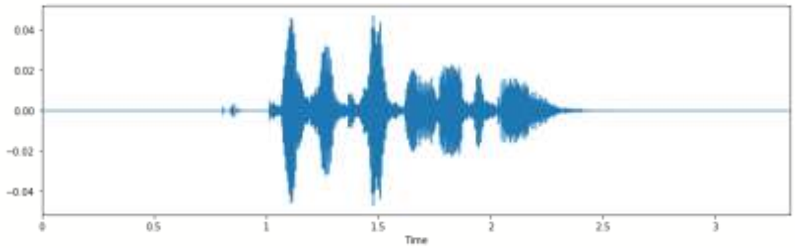
The waveplot for audio with happy emotion ranges from 1 to 2.5 at a given time.



The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

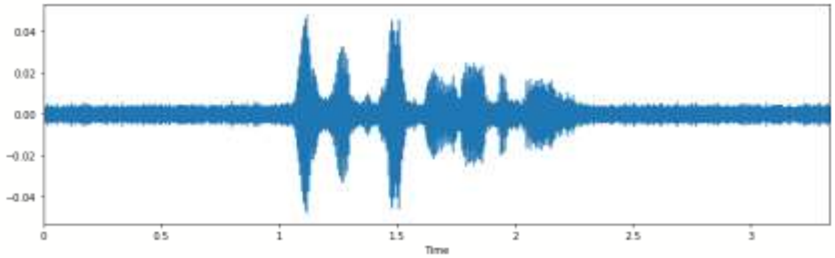


Let's take any example and checking for techniques.

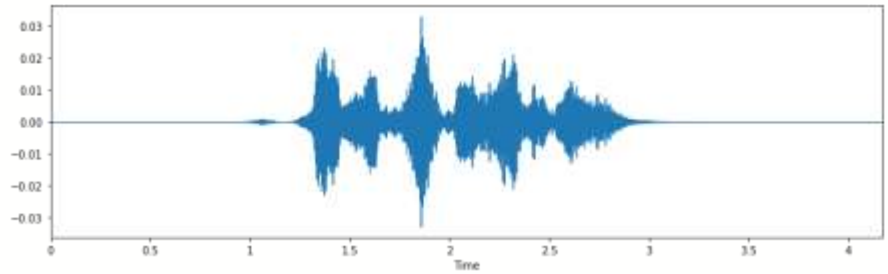


x = noise(data)

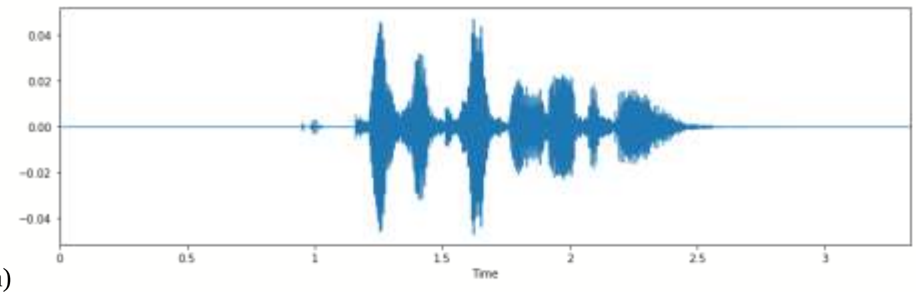
Noise injection is a very good augmentation technique because of which we can assure our training model is not overfitted.



x = stretch(data)



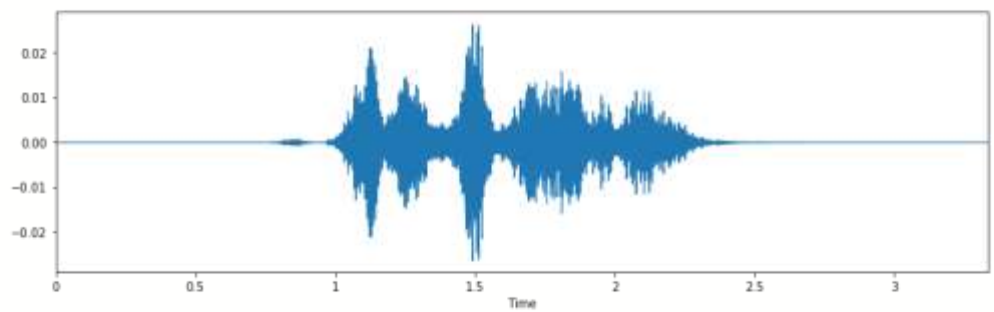
The above shows the stretch technique for the given data.



x = shift(data)

The above shows the shift technique for the given data.

x = pitch(data, sample_rate)

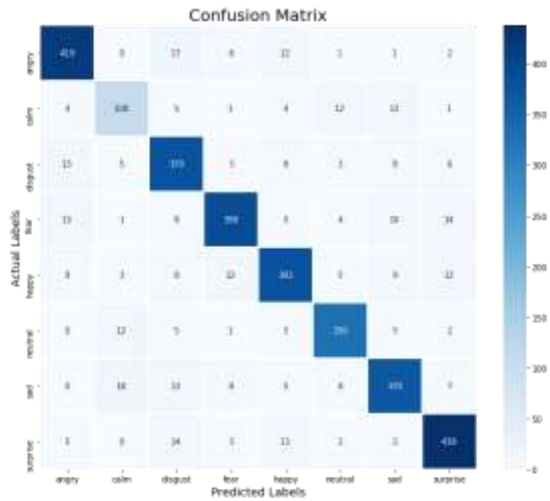


The above shows the pitch technique for the given data.

Predicted vs Actual

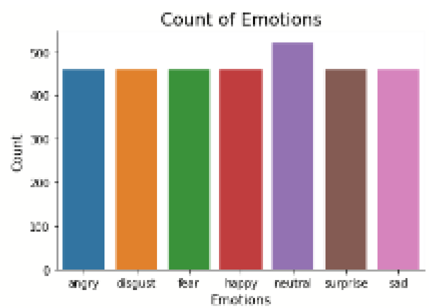
	Predicted Labels	Actual Labels
0	happy	happy
1	fear	fear
2	calm	calm
3	neutral	neutral
4	fear	fear
5	fear	fear
6	happy	happy
7	surprise	surprise
8	happy	happy
9	happy	happy

This is the confusion matrix

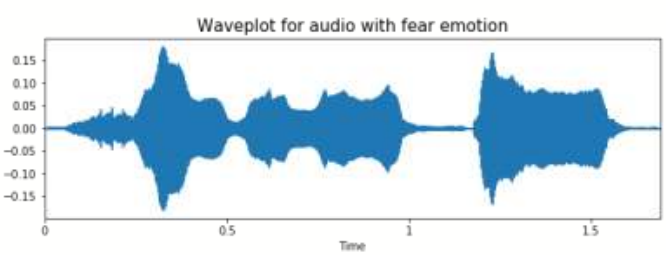


TESS+SAVEE

The below bar graph shows the count of emotions. Angry, Disgust, Fear, Happy, Surprise, and Sad are in between 400-500. For Neutral, it is 500.

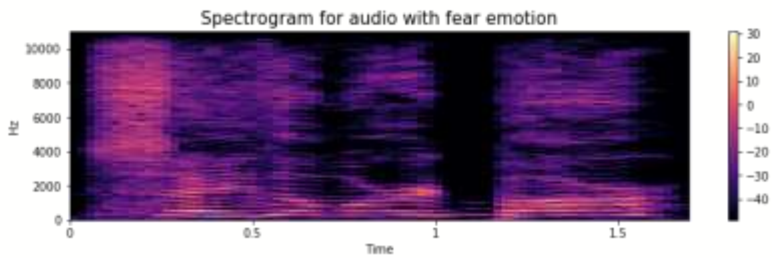


The waveplot for audio with fear emotion ranges from 0 to 1.5 at a given time.

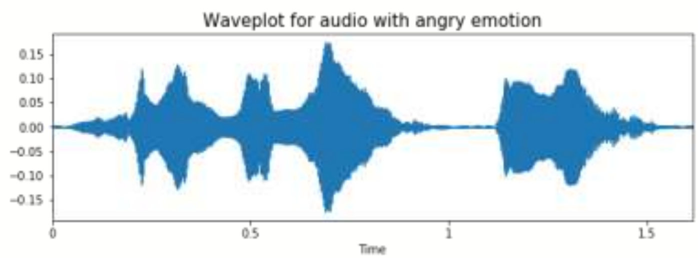


International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

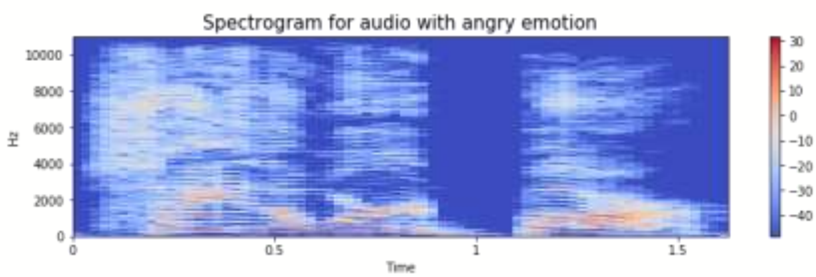
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



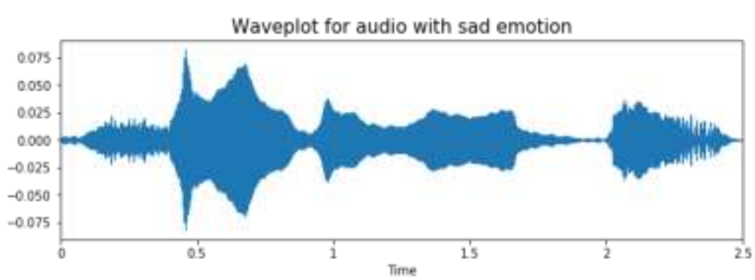
The waveplot for audio with angry emotion ranges from 0 to 1.5 at a given time.



The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

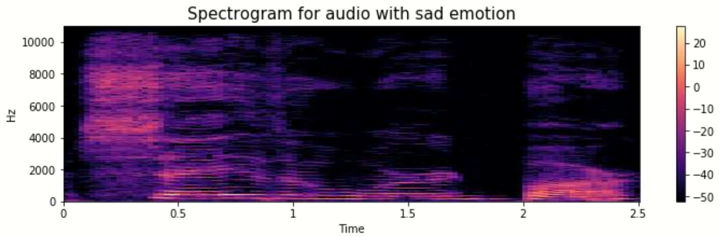


The waveplot for audio with sad emotion ranges from 0 to 2.5 at a given time.

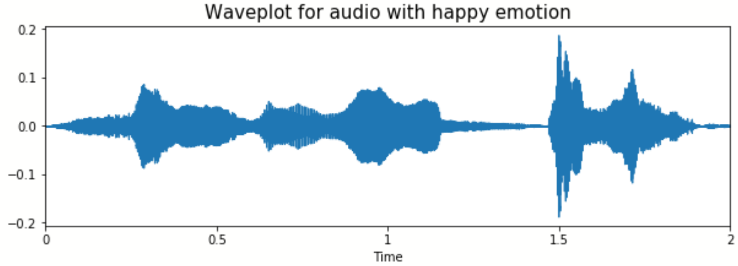


The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

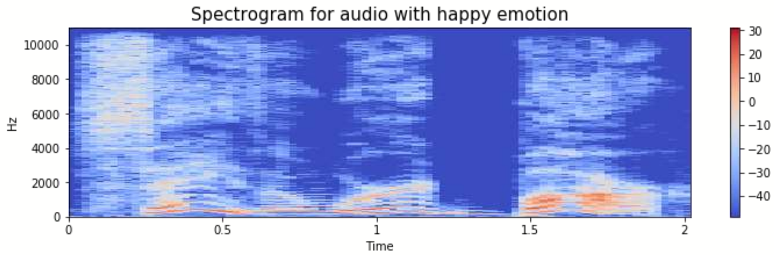
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



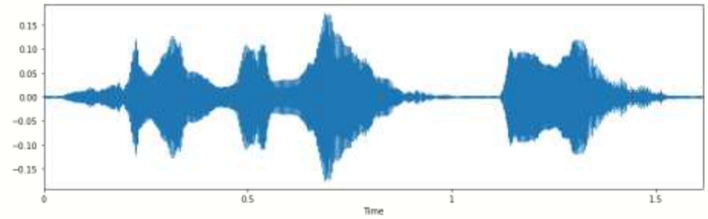
The waveplot for audio with happy emotion ranges from 0 to 2 at a given time.



The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

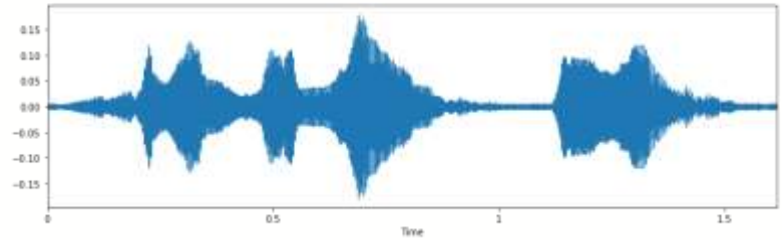


Let's take any example and checking for techniques.

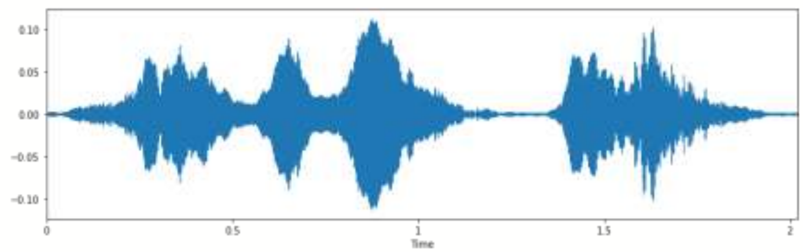


Noise injection is a very good augmentation technique because of which we can assure our training model is not overfitted.

x = noise(data)

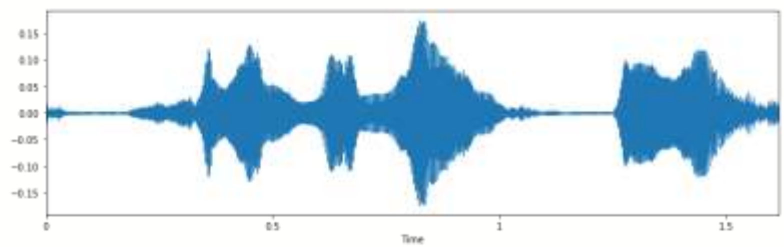


x = stretch(data)



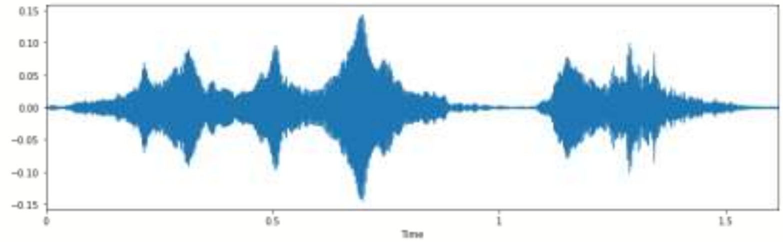
The above shows the stretch technique for the given data.

x = shift(data)



The above shows the shift technique for the given data.

x = pitch(data, sample_rate)

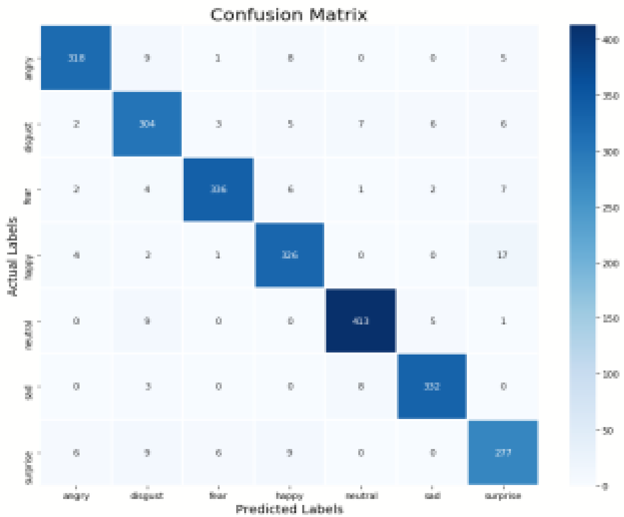


The above shows the pitch technique for the given data.

Predicted vs Actual

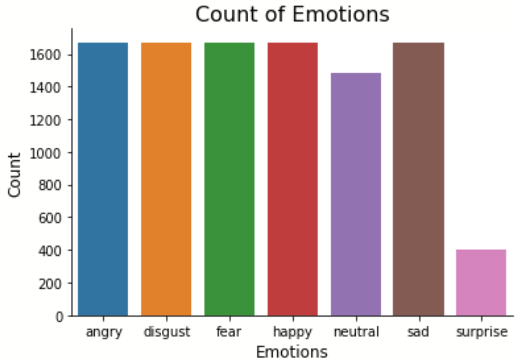
	Predicted Labels	Actual Labels
0	happy	happy
1	sad	sad
2	disgust	disgust
3	sad	sad
4	neutral	neutral
5	happy	happy
6	happy	happy
7	angry	angry
8	disgust	surprise
9	neutral	neutral

This is the confusion matrix

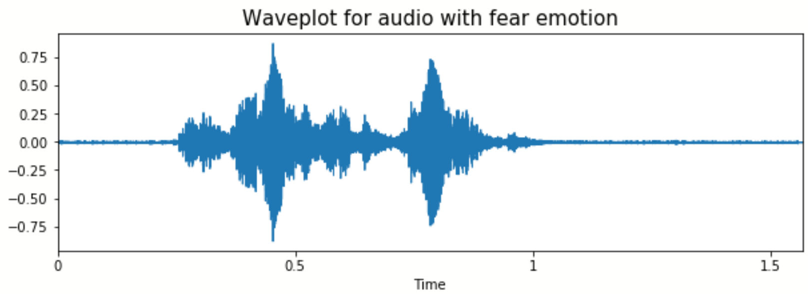


TESS+CREMA

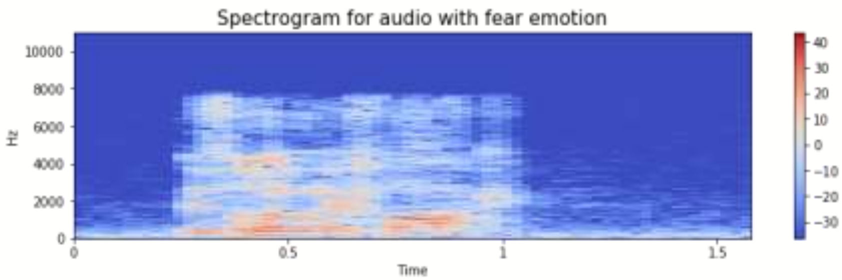
The below bar graph shows the count of emotions. Angry, Disgust, Fear, Happy, and Sad are above 600. For Neutral, it is in between 1400-1600, and for surprise, it is 400.



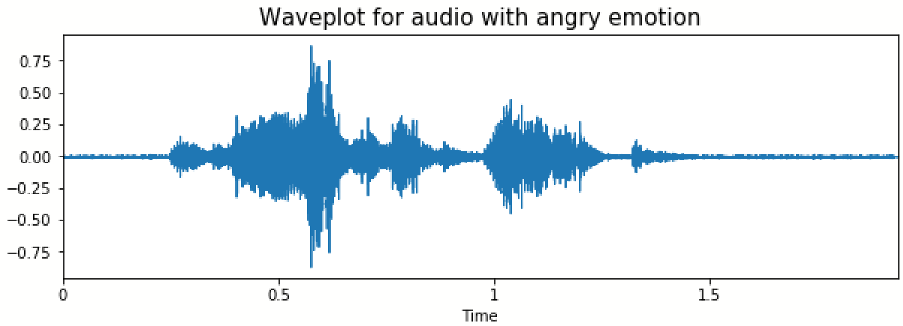
The waveplot for audio with fear emotion ranges from 0 to 1 at a given time.



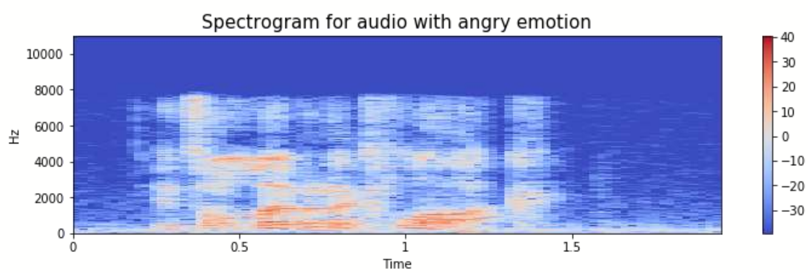
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



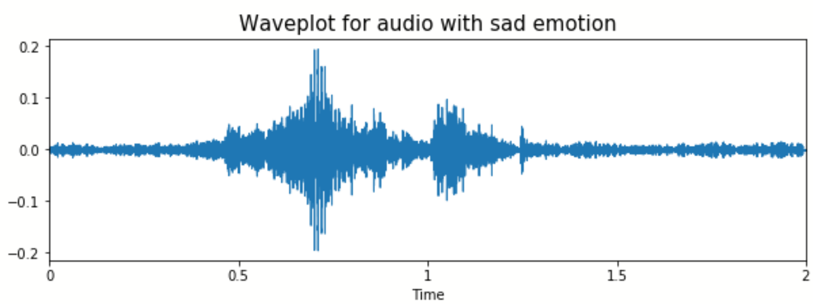
The waveplot for audio with angry emotion ranges from 0 to 1.5 at a given time.



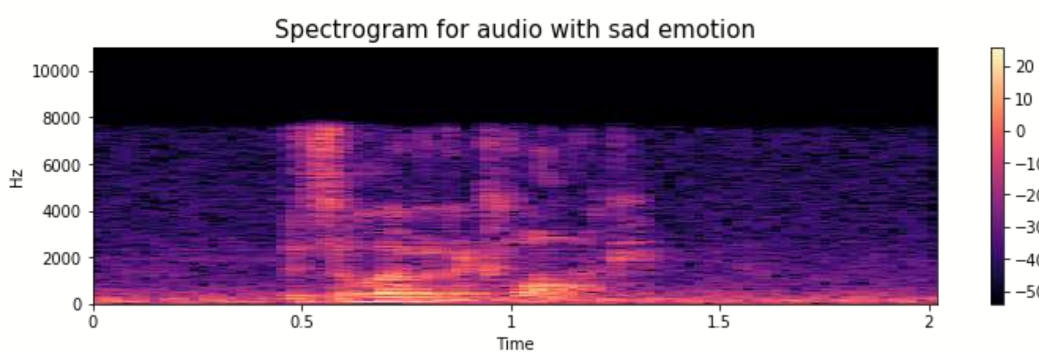
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



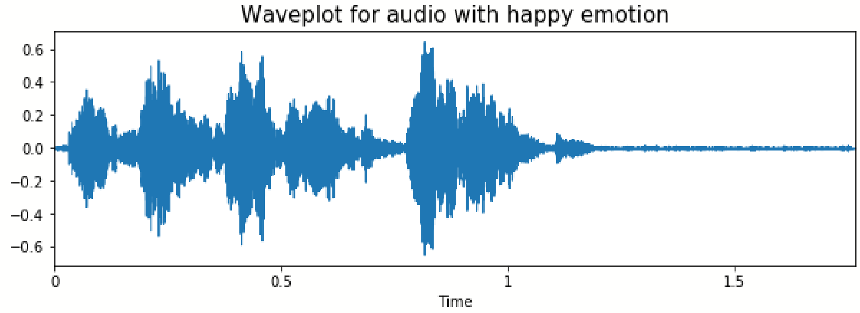
The waveplot for audio with sad emotion ranges from 0 to 1.5 at a given time.



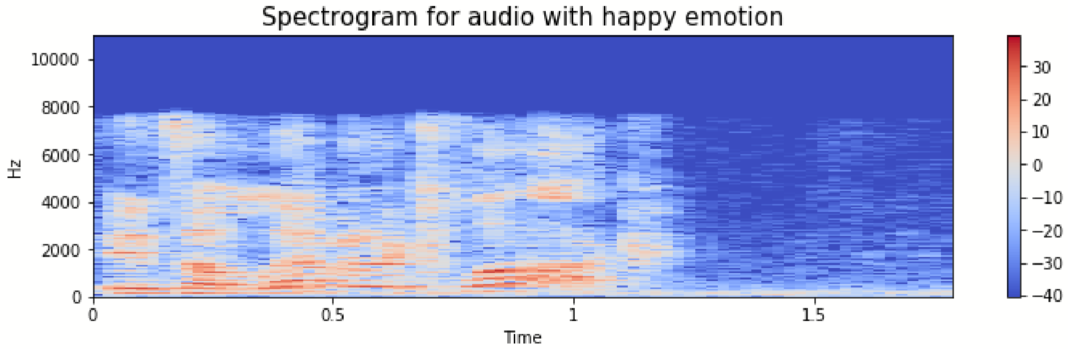
The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.



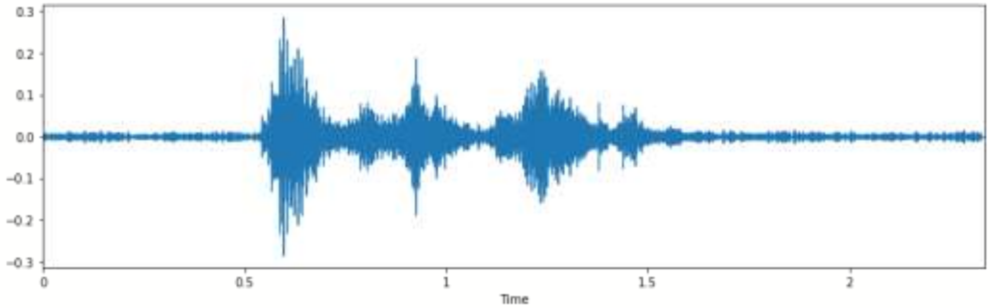
The waveplot for audio with happy emotion ranges from 0 to 1 at a given time.



The spectrogram shows the visual representation of frequencies changing with respect to time for given audio signals.

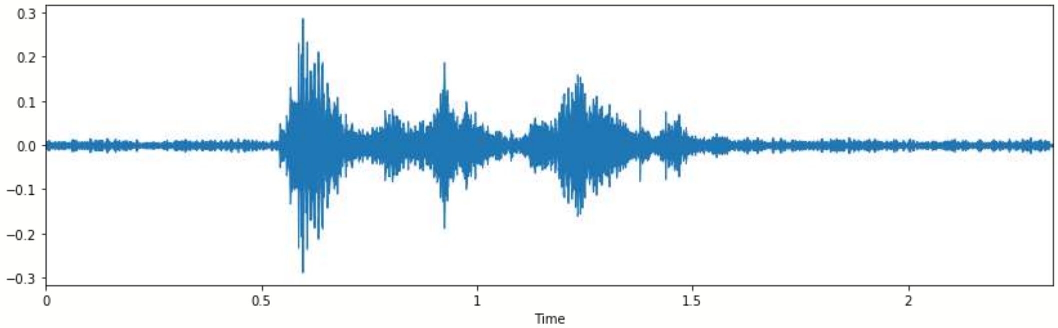


Let's take any example and checking for techniques.

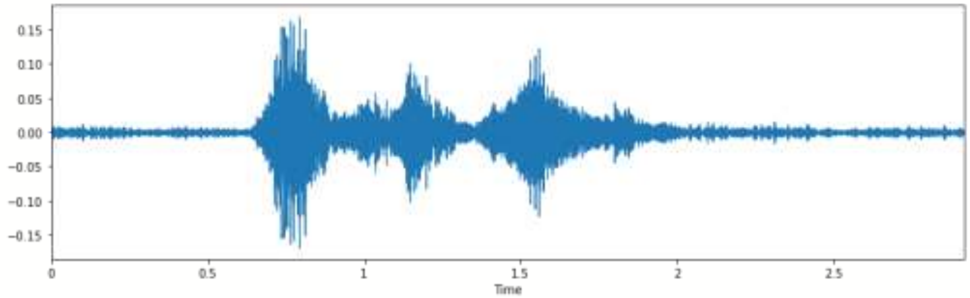


Noise injection is a very good augmentation technique because of which we can assure our training model is not overfitted.

x = noise(data)

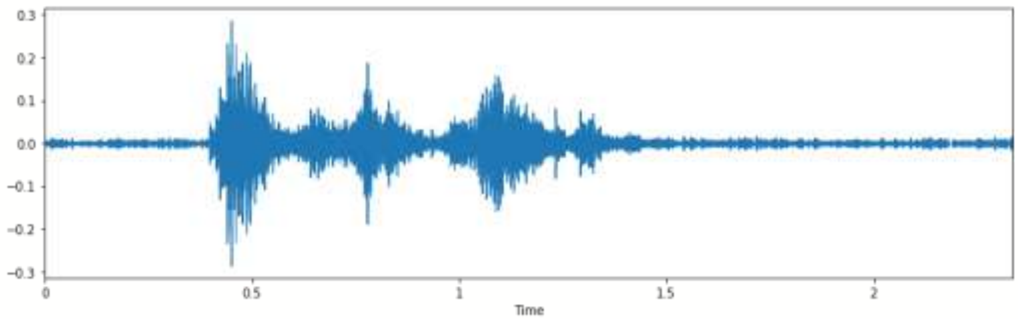


x = stretch(data)



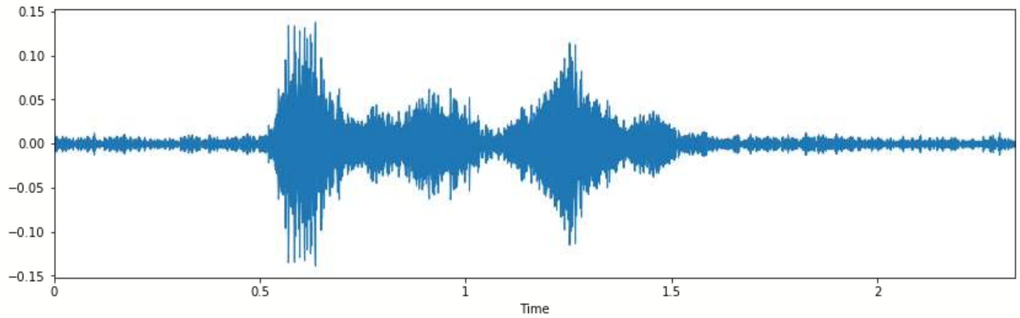
The above shows the stretch technique for the given data.

x = shift(data)



The above shows the shift technique for the given data.

x = pitch(data, sample_rate)



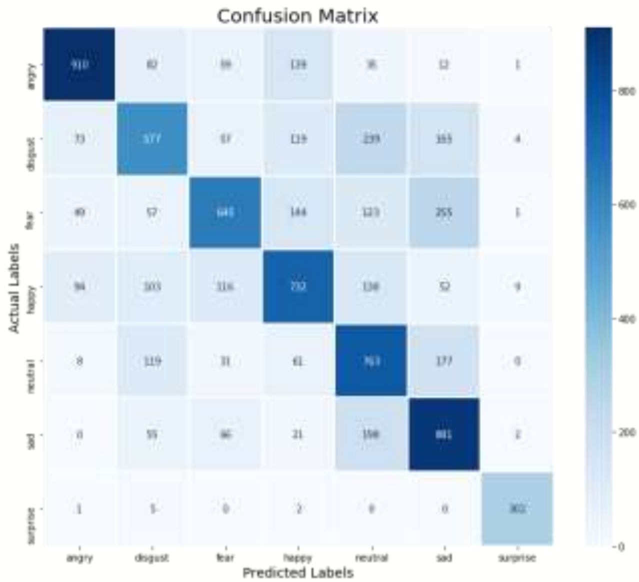
The above shows the pitch technique for the given data.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Predicted vs Actual

	Predicted Labels	Actual Labels
0	disgust	neutral
1	happy	disgust
2	happy	happy
3	disgust	fear
4	happy	angry
5	happy	neutral
6	angry	angry
7	angry	angry
8	neutral	neutral
9	neutral	neutral

This is the confusion matrix



5. Result

1) RAVDESS+CREMA+SAVEE+TESS

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

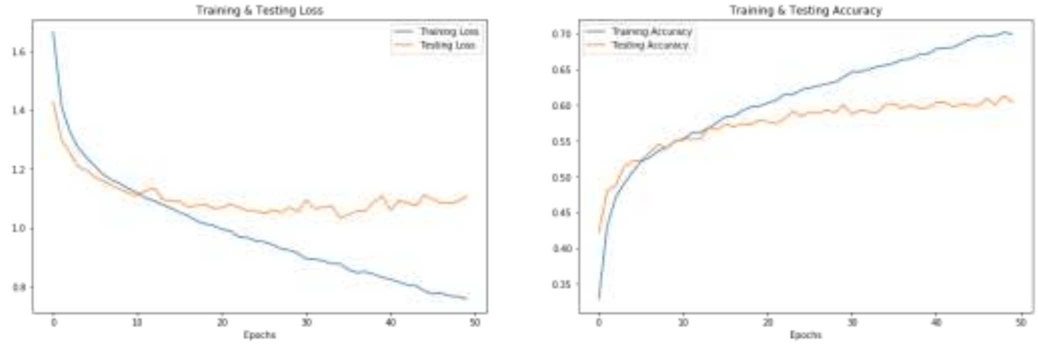


Fig5.1: Training and Testing Loss and Accuracy of all the data sets

	precision	recall	f1-score	support
angry	0.76	0.71	0.74	1438
calm	0.61	0.77	0.68	137
disgust	0.55	0.48	0.51	1468
fear	0.69	0.45	0.55	1424
happy	0.51	0.60	0.55	1462
neutral	0.54	0.65	0.59	1310
sad	0.58	0.64	0.61	1400
surprise	0.78	0.84	0.81	483
accuracy			0.60	9122
macro avg	0.63	0.64	0.63	9122
weighted avg	0.61	0.60	0.60	9122

Fig5.1.1: Classification report of the model with all datasets

2) RAVDESS

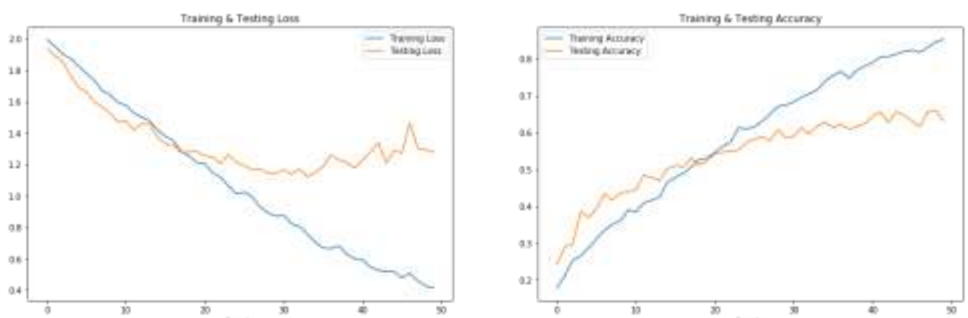


Fig 5.2: Training and Testing Loss and Accuracy of Ravdess dataset

International Journal for Innovative
Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

	precision	recall	f1-score	support
angry	0.72	0.67	0.70	132
calm	0.76	0.77	0.76	159
disgust	0.56	0.61	0.58	135
fear	0.61	0.66	0.64	151
happy	0.49	0.54	0.51	141
neutral	0.41	0.40	0.41	72
sad	0.72	0.52	0.60	151
surprise	0.70	0.77	0.74	139
accuracy			0.63	1080
macro avg	0.62	0.62	0.62	1080
weighted avg	0.64	0.63	0.63	1080

Fig 5.2.1: Classification report of the model with Ravdess dataset

3) CREMA

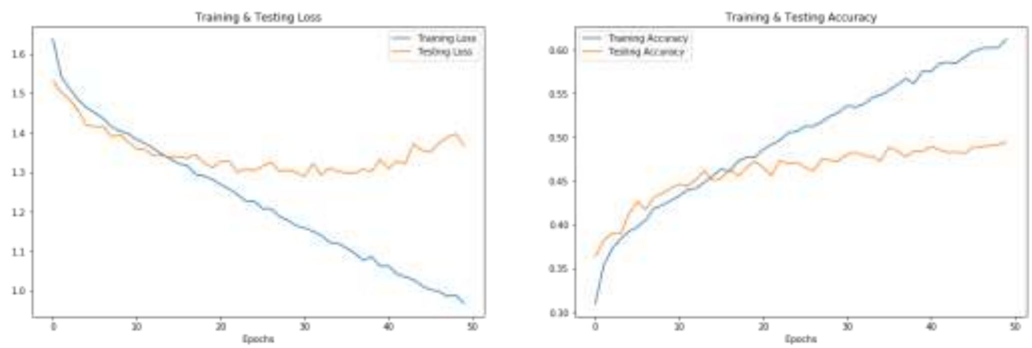


Fig 5.3: Training and Testing Loss and Accuracy of CREMA dataset

	precision	recall	f1-score	support
angry	0.73	0.64	0.68	938
disgust	0.41	0.28	0.33	934
fear	0.53	0.38	0.44	960
happy	0.49	0.49	0.49	967
neutral	0.41	0.50	0.45	857
sad	0.45	0.68	0.54	926
accuracy			0.49	5582
macro avg	0.50	0.50	0.49	5582
weighted avg	0.50	0.49	0.49	5582

Fig5.3.1: Classification report of the model with CREMA dataset

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

4) SAVEE

Fig5.4: Training and Testing Loss and Accuracy of SAVEE dataset

	precision	rec
angry	0.68	0
disgust	0.49	0
fear	0.52	0
happy	0.55	0
neutral	0.77	0
sad	0.63	0
surprise	0.45	0
accuracy		
macro avg	0.58	0
weighted avg	0.61	0

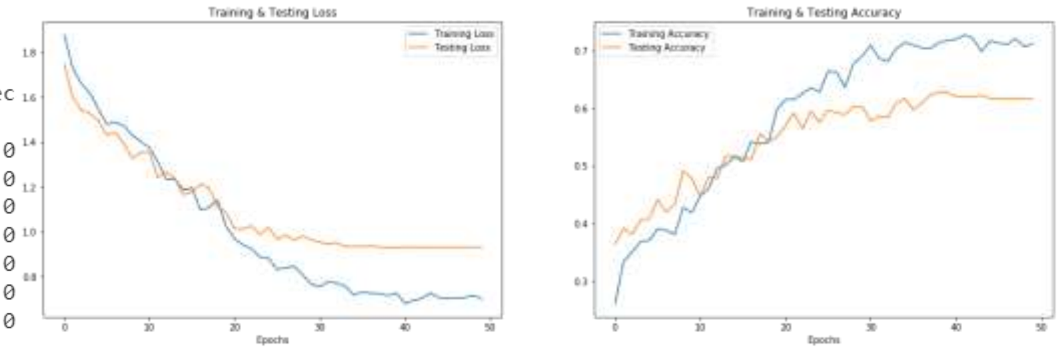


Fig 5.4.1: Classification report of the model with SAVEE dataset

5) TESS

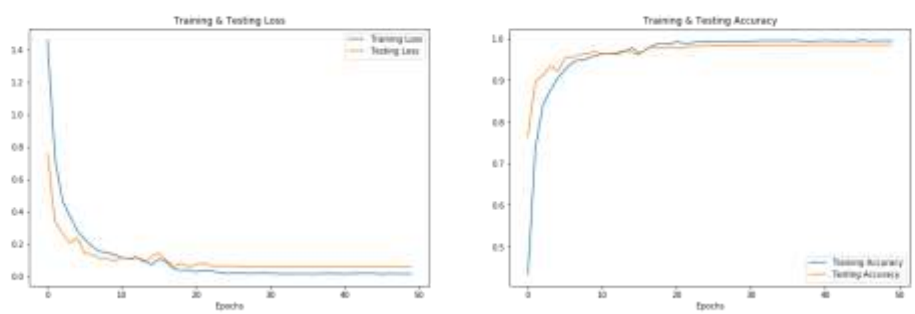


Fig5.5: Training and Testing Loss and Accuracy of TESS dataset

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

	precision	recall	f1-score	support
angry	1.00	0.99	0.99	308
disgust	0.97	0.98	0.97	291
fear	0.99	1.00	1.00	303
happy	0.99	0.98	0.99	310
neutral	1.00	1.00	1.00	322
sad	0.99	1.00	1.00	279
surprise	0.95	0.96	0.95	287
accuracy			0.99	2100
macro avg	0.98	0.99	0.98	2100
weighted avg	0.99	0.99	0.99	2100

Fig 5.5.1: Classification report of the model with TESS dataset

6) RAVDESS+CREMA

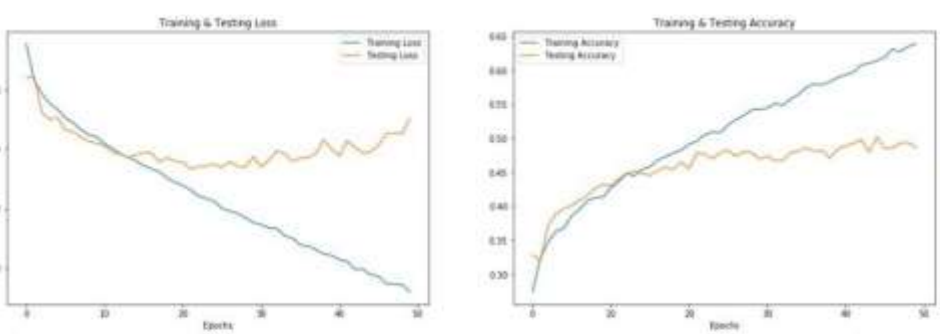


Fig5.6: Training and Testing Loss and Accuracy of RAVDESS+CREMA dataset

	precision	recall	f1-score	support
angry	0.64	0.71	0.67	1094
calm	0.63	0.76	0.69	144
disgust	0.39	0.41	0.40	1066
fear	0.60	0.26	0.37	1121
happy	0.41	0.47	0.44	1062
neutral	0.39	0.52	0.44	910
sad	0.53	0.49	0.51	1129
surprise	0.60	0.65	0.63	136
accuracy			0.49	6662
macro avg	0.52	0.54	0.52	6662
weighted avg	0.50	0.49	0.48	6662

Fig 5.6.1: Classification report of the model with RAVDESS+CREMA dataset

7) RAVDESS+ SAVEE

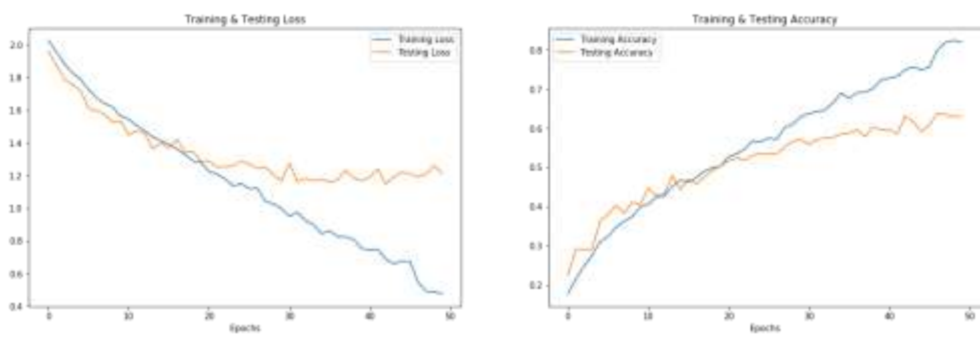


Fig5.7: Training and Testing Loss and Accuracy of RAVDESS and SAVEE dataset

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

	precision	recall	f1-score	support
angry	0.67	0.74	0.71	188
calm	0.67	0.73	0.70	150
disgust	0.61	0.52	0.56	184
fear	0.60	0.60	0.60	193
happy	0.58	0.52	0.55	187
neutral	0.66	0.68	0.67	161
sad	0.55	0.54	0.55	199
surprise	0.68	0.74	0.71	178
accuracy			0.63	1440
macro avg	0.63	0.63	0.63	1440
weighted avg	0.63	0.63	0.63	1440

Fig5.7.1: Classification report of the model with RAVDESS and SAVEE dataset

8) RAVDESS+ TESS

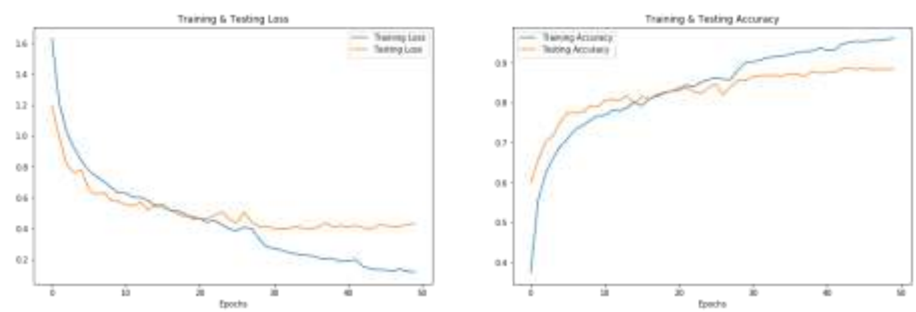


Fig5.8: Training and Testing Loss and Accuracy of RAVDESS and TESS dataset

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

	precision	recall	f1-score	support
angry	0.91	0.91	0.91	458
calm	0.73	0.73	0.73	148
disgust	0.84	0.89	0.87	425
fear	0.92	0.88	0.90	445
happy	0.88	0.87	0.87	438
neutral	0.90	0.92	0.91	360
sad	0.89	0.86	0.87	429
surprise	0.91	0.92	0.91	477
accuracy			0.89	3180
macro avg	0.87	0.87	0.87	3180
weighted avg	0.89	0.89	0.89	3180

Fig5.8.1: Classification report of the model with RAVDESS and TESS dataset

9) TESS+SAVEE

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

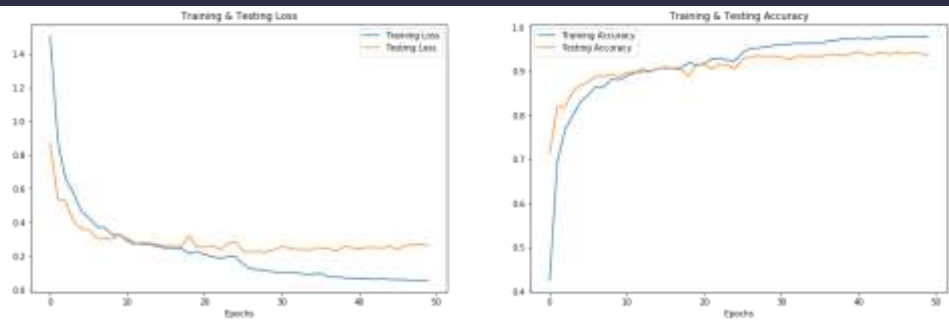
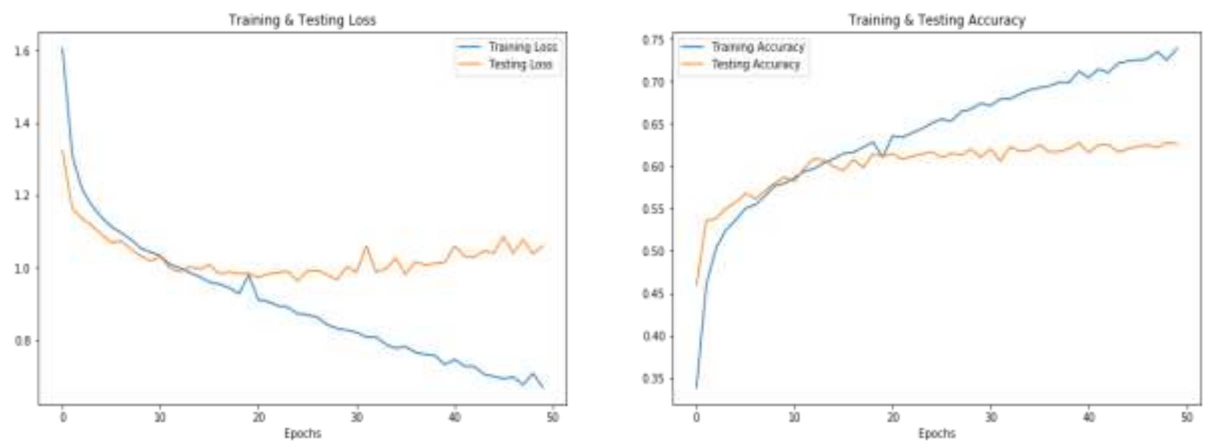


Fig5.9: Training and Testing Loss and Accuracy of SAVEE and TESS

	precision	recall	f1-score	support
angry	0.96	0.93	0.95	341
disgust	0.89	0.91	0.90	333
fear	0.97	0.94	0.95	358
happy	0.92	0.93	0.93	350
neutral	0.96	0.96	0.96	428
sad	0.96	0.97	0.97	343
surprise	0.88	0.90	0.89	307
accuracy			0.94	2460
macro avg	0.94	0.94	0.94	2460
weighted avg	0.94	0.94	0.94	2460

Fig5.9.1: Classification report of the model with SAVEE and TESS



10) TESS+CREMA

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Fig 5.10: Training and Testing Loss and Accuracy of CREMA and TESS dataset

	precision	recall	f1-score	support
angry	0.80	0.74	0.77	1238
disgust	0.58	0.47	0.52	1234
fear	0.66	0.51	0.57	1274
happy	0.60	0.59	0.59	1244
neutral	0.51	0.66	0.57	1159
sad	0.57	0.72	0.64	1223
surprise	0.95	0.97	0.96	310
accuracy			0.63	7682
macro avg	0.67	0.66	0.66	7682
weighted avg	0.64	0.63	0.63	7682

Fig 5.10.1: Classification report of the model with CREMA and TESS dataset

As we observed from the above graphs Tess dataset has given the highest accuracy of 98.52%. This dataset when combined with other datasets has increased the accuracy of the model than using the other datasets individually.

For example, RAVDESS dataset model gave an accuracy of 63.24 but when combined with TESS it has given a good accuracy of 88.52.

This implies that TESS is a preferred dataset than the other dataset by seeing how it is improving the accuracy of the model and also by giving highest accuracy even when it is individually. If someone wants to use any particular dataset it is better to add tess to it and we should also make sure that there is a bit of balance of male and female voices after combining the datasets.

6. Conclusion

Overall achieved 60.4% accuracy on RAVDESS + CREMA + SAVEE + TESS data , 63.24% with RAVDESS, 49.46% with CREMA, 61.66% with SAVEE, 98.52% with TESS, 48.63% with RAVDESS+CREMA, 62.91% with RAVDESS+SAVEE, 88.52% with RAVDESS+TESS, 62.61% with CREMA + TESS, 93.73% with SAVEE+TESS. As we observed from the above graphs Tess dataset has given the highest accuracy of 98.52%. This dataset when combined with other datasets has increased the accuracy of the model than using the other datasets individually.

This implies that TESS is a preferred dataset than the other dataset by seeing how it is improving the accuracy of the model and also by giving highest accuracy even when it is individually. If someone wants to use any particular dataset it is better to add tess to it and we should also make sure that there is a bit of balance of male and female voices after combining the datasets.

7. Future Scope

Emotion recognition has wide scope in many areas such as human computer interaction, biometric security etc. So it provides insight into artificial intelligence or machine intelligence that uses various supervised and unsupervised machine-learning algorithms to simulate the human brain. Emotion recognition plays a crucial role in the era of Artificial intelligence and Internet of things. It offers tremendous scope to human computer interaction, robotics, health care, biometric security and behavioral modeling. Emotion recognition systems recognize emotions from facial expressions, text data, body movements, voice, brain or heart signals. Emotion recognition has wide scope in many areas such as human computer interaction, biometric security etc. So it provides insight into artificial intelligence or machine intelligence that uses various supervised and unsupervised machine-learning algorithms to simulate the human brain.

8. References

1. https://scholar.google.com/scholar?as_ylo=2018&q=Speech+Based+Human+Emotion+Recognition+Springer+speech+technology+S.+Sangeeta+Sastra+University&hl=en&as_sdt=0,5&as_vis=1#d=gs_qabs&u=%23p%3D2QjUkaAWi_cJ
2. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Speech+Based+Human+Emotion+Recognition+S.+Jothilakshmi+annamalai+University+&btnG=#d=gs_qabs&u=%23p%3DwdD8fMFQk8AJ
3. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Speech+Based+Human+Emotion+Recognition+S.+Jothilakshmi+annamalai+University+&btnG=#d=gs_qabs&u=%23p%3DnkeGA9CuSO0J
4. https://en.wikipedia.org/wiki/Audio_signal
5. https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623
6. https://pnsn.org/spectrograms/
7. https://www.kaggle.com/
8. Anisha P R , Kishor Kumar Reddy C and Nguyen Gia Nhu, “Blockchain Technology: A Boon at the Pandemic Times – A Solution for Global Economy Upliftment with AI and IoT”, EAI/Springer Innovations in Communication and Computing, 2022.

9. PR Anisha, CKK Reddy, NG Nhu, Blockchain Technology: A Boon at the Pandemic Times–A Solution for Global Economy Upliftment with AI and IoT, Blockchain Security in Cloud Computing, 227-252, 2022
10. PR Anisha, Kishor Kumar Reddy C, NG Nguyen, G Sreelatha, A Text Mining using Web Scraping for Meaningful Insights, Journal of Physics: Conference Series 2089 (1), 012048, 2021
11. CKK Reddy, PR Anisha, RM Mohana, Assessing Wear Out of Tyre using Opencv & Convolutional Neural Networks, Journal of Physics: Conference Series 2089 (1), 012001, 2021
12. T Lingala, CKK Reddy, BVR Murthy, R Shastry, Y Pragathi, L-Diversity for Data Analysis: Data Swapping with Customized Clustering, Journal of Physics: Conference Series 2089 (1), 012050, 2021
13. SN Paidimarry, CKK Reddy, LG Lolla, Internet of things enabled smart baggage follower with theft prevention, Journal of Physics: Conference Series 1950 (1), 012002, 2021
14. PR Anisha, CKK Reddy, K Apoorva, CM Mangipudi, Early Diagnosis of Breast Cancer Prediction using Random Forest Classifier, IOP Conference Series: Materials Science and Engineering 1116 (1), 012187, 2021
15. RM Mohana, CKK Reddy, PR Anisha, BVR Murthy, Random forest algorithms for the classification of tree-based ensemble, Materials Today: Proceedings, 2021
16. Kishor Kumar Reddy C, Anisha P R, Shastry R, Ramana Murthy B V, "Comparative Study on Internet of Things: Enablers and Constraints", Advances in Intelligent Systems and Computing, 2021
17. Kishor Kumar Reddy C, Anisha P R, Apoorva K, "Early Prediction of Pneumonia using Convolutional Neural Network and X-Ray Images", Smart Innovation, Systems and Technologies, 2021
18. R Madana Mohana, Kishor Kumar Reddy C and Anisha P R, "A Study and Early Identification of Leaf Diseases in Plants using Convolutional Neural Network", Springer 4th Int Conference on Smart Computing and Informatics, 2020, India
19. Anisha P R, C Kishor Kumar Reddy and Nuzhat Yasmeen, "Predicting the Energy Output of Wind Turbine Based on Weather Condition", Springer 4th Int Conference on Smart Computing and Informatics, 2020, India