1- INTRODUCTION:

I have selected this project because Audio Classification domain is growing in various fields like NLP, Voice Recognition, chatbots etc. Dataset which is used is UrbanSound8k, which is available in kaggle as well as on other websites. This dataset contains 8732 labeled sound from 10 classes i.e air_conditioner, car_horn, children_playing, dog_bark, drilling, enginge_idling, gun_shot, jackhammer, siren, and street_music.An addition to this, file also contain a csv file , which have information about all sounds.

2- OBJECTIVES:

The objective of this is to predict the accuracy using different machine learning algorithm and ANN

With different features of Sound.

3- METHODOLOGY:

- Loading of Audio with labels.
- Audio Augmentation(balancing) for unbalanced data.
- Feature Extraction.
- Splitting of Data as test and train
- Predicting and checking accuracy by using different ML algorithm.

3.1- Loading of Audio With Labels.

Librosa library was used to load the audio.

3.2- Audio Augmentation(balancing) for unbalanced data.

There are various method for audio augmentation, out which noise addition, time stretching and pitch addition was used. Librosa and numpy library was used for augmentation.

(Note: Python library *torch_audiomentations* can be used for various audio augmentation.)

3.2- Feature Extraction.

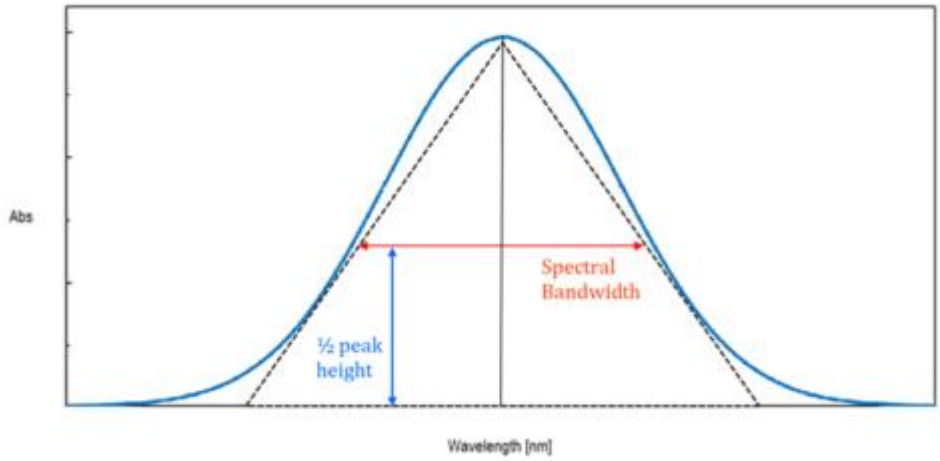
For feature extraction I have used librosa.feature.

Zero_crossing_rate: The zero-crossing rate (ZCR) is the rate at which a signal changes from positive to zero to negative or from negative to zero to positive.

Root_mean_square: RMS, which stands for root mean square, is a metering tool that measures the average loudness of an audio track . The value displayed is an average of the audio signal. The RMS value will give you a more accurate look at the perceived loudness of the music track for the average listener

Spectral_centroids: This is the most commonly used audio feature. It gives the center of gravity of the magnitude spectrum or, in other words, it indicates the central mass of the spectrum. It basically gives the frequency band where most of the energy is concentrated.

Spectral bandwidth: Bandwidth is the difference between the upper and lower frequencies in a continuous band of frequencies. As we know the signals oscillate about a point so if the point is the centroid of the signal then the sum of maximum deviation of the signal on both sides of the point can be considered as the bandwidth of the signal at that time frame.



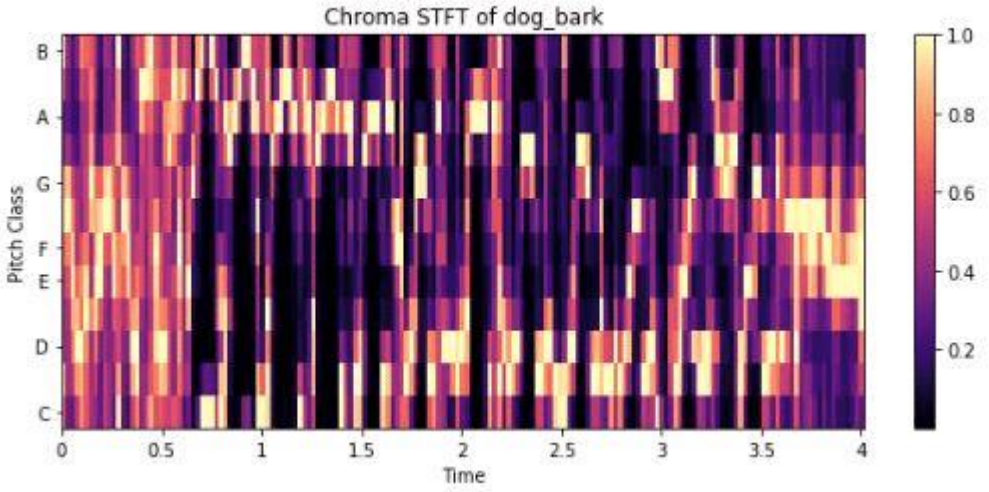
Mel-Frequency Cepstral Coefficients (MFCCs): Short term power spectrum of any sound represented by the Mel frequency cepstral (MFC) and combination of MFCC makes the MFC. It can be derived from a type of inverse Fourier transform(cepstral) representation. MFC allows a better representation of sound because in MFC the frequency bands are equally distributed on the Mel scale which approximates the human auditory system's response more closely.

It can be derived by mapping the Fourier transformed signal onto the male scale using triangle or cosine overlapping windows. Where after taking the logs of the powers at each of the Mel frequencies and after discrete cosine transform of the Mel log powers give the amplitude of a spectrum. The amplitude list is MFCC.

Magnitude spectrum: The Magnitude Spectrum of a signal describes a signal using frequency and amplitude. That is frequency components of a periodic signal are plotted using Frequency Domain - frequencies plotted in X-axis and amplitude plotted in Y-axis**.**

Mel spectrogram: Mel scale is the scale of pitches that can be felt by the listener to be equal in distance from one another. For example, a listener can identify the difference between the audio of 10000 Hz and 15000 Hz if the audio sources are in the same distance and atmosphere. Representation of frequencies into the Mel scale generates the Mel spectrogram. Frequencies can be converted into the Mel scale using the Fourier transform.

Chroma STFT: Chroma STFT The Chroma value of an audio basically represent the intensity of the twelve distinctive pitch classes that are used to study music. They can be employed in the differentiation of the pitch class profiles between audio signals. Chroma STFT used short-term Fourier transformation to compute Chroma features. STFT represents information about the classification of pitch and signal structure. It depicts the spike with high values (as evident from the color bar net to the graph) in low values (dark regions).



Spectral Contrast: In an audio signal, the spectral contrast is the measure of the energy of frequency at each timestamp. Since most of the audio files contain the frequency whose energy is changing with time. It becomes difficult to measure the level of energy. Spectral contrast is a way to measure that energy variation.

Spectral flatness: The spectral flatness (also called Wiener entropy) is defined as the ratio of the geometric mean of a spectrum to its arithmetic mean.

4-Model Building With Machine Learning Algorithm and ANN:

K nearest neighbors, Random Forest, Extra Tree Classifier, Support Vector Classifier, XGboost Classifier, Voting Classifier, Stacking Classifier

5-Conclusion:

Deep Learning Algorithm is best suited for Sound Classification.

(Audio Data can be downloaded from Kaggle)