CLASSIFICATION OF ENVIRONMENTAL SOUNDS USING DEEP LEARNING

R.Gayathri[1], B.Pavithra[2], S.Prabhavathi[3], G. Sakthi[4]

[1]Assistant Professor, Dept. of Electronics and Communication Engineering, Tamil Nadu, India
[2]UG Student, Dept. of Electronics and Communication Engineering, Tamil Nadu, India
[3]UG Student, Dept. of Electronics and Communication Engineering, Tamil Nadu, India
[4]UG Student, Dept. of Electronics and Communication Engineering, Tamil Nadu, India

---***---

Abstract - *One of the most common used applications in Audio Deep Learning/Machine Learning is Sound Classification. It involves learning to classify sounds and to predict the category of that specific sound. Aim of our project is to classify the environmental audio sound using deep learning Techniques. The Urban sound8K dataset audio samples is downloaded and were considered for our project work. Our project work implemented in three categories. Firstly, the features of audio signals are extracted using MFCC's and the sample rate of each audio is predicted using librosa and scipy python package. In the second part, the skeletal format for ANN is built and were trained for the extracted feature of audio signal. The training process is repeated for various iteration with respect to achieve better accuracy in third part. Our project concepts applied to many practical scenarios e.g. to identify and classify the different music clips using this models or classifying the voice from speakers.*

Key Words: Audio signal, machine learning, Deep Learning Technique, Python Package

1. INTRODUCTION

Sound recognition is a way of analyzing audio signals that is based on classic pattern recognition technology.. Data processing, feature extraction and classification algorithms from sound recognition technologies. Sound recognition can categorize feature vectors. Preliminary data processing and linear predictive coding comes under the results from Feature vectors

Sound recognition technologies are used for:

1. Recognition of music
2. Recognition of speech

For surveillance and monitoring systems, the acoustic environment is based on automatic alarm detection and identification. Assistance to disabled person or elderly people affected in their person hearing capabilities. Identifying species of animals like fish and mammals, e.g. in acoustical oceanography.

2. Body of Paper

This paper brings up Audio signals are important part in any kind of communication media today, where the speech recognition helps to understand and visualize the various scenario. This paper deals with audio classification and the diverse content it contains. The recognition using the approaches such as classification, regression and other approaches work with identification of such analysis is talked over. This paper comprises recognition accuracy is Audio recognition and its analysis tool work with the fundamental of working procedure. HMM Recognition is a pattern matching and analysis process work with the HMM recognition while it deals the recognition of signal based on HMM training. Speech breakup, Noise cancellation and processing of actual data work with help of pre-processing unit. They have inspected 13 highlights suggest to quantify thoughtfully particular properties of discourse and additionally music flag, and joined them in a few multidimensional classifying structures. For the datasets right now being used, the best classifier arranges with 5.8% blunder on a casing by-outline premise, and 1.4% mistake when coordinating long (2.4 second) fragments of audio. This paper designed as Sound is the signal form which is always around us and many of the latest technology, applications are working on the sound system. Finding an appropriate technique to do a better categorization is always a difficult issue when working with a dynamic operation. The presented approach which is SENet is given which the computing oriented fast approach for more accuracy and less error rate. This approach is implemented using the MATLAB tool and further the parameter such as Accuracy and Error rate is computed using classification confusion matrix. The output of this paper is the high value of accuracy and less value of error rate.

This paper offers a new music classification model based on metric learning, with the primary goal of learning a unique measure for each customer. The learning of a group of parameterized distances employing a structured prediction approach from a set of MP3 audio files containing several music genres according to the users decision They used the Mel-Frequency Cepstral Coefficient (MFCC) to extract the auditory data and Principal Components Analysis to reduce the dimensionality (PCA).The model validity performing a group of experiments and comparing the training and testing outputs with baseline algorithms, such as K-means and Soft Margin Linear Support Vector Machine (SVM). The model presented a stable performance even when reducing the training set size and the audio segment length. The results obtained with the GTZAN dataset are congruent with those found in the literature, with most of the referred papers showing that the GTZAN dataset is superior. The performance using 50% of the metric learning model used and the data for training and 50% for testing indicates that, with the increment in the number of training constraints, the model tends to better evolve its generalization power when compared to SVM and others referenced grouping. They also demonstrate prediction capability by obtaining good prediction results on the GTZAN dataset.

The literature reveals that the study frequently utilize western or western classical music whereas no significant short is reported in computational modeling of cultural-specific music e.g., Sri Lankan folk melodies, despite of being an abundant source of emotion expression. Further, the applicability of existing division trained using different ground-truth data, in other cultural-specific content is found to be problematic. As a result, they used machine learning techniques to classify emotions in a dataset of Sri Lankan folk songs. Acoustic features relevant to dynamics, rhythm, timbre, pitch, and tone were retrieved using the MATLAB MIR Toolbox Five common classification algorithms (Support Vector Machine, Naive Bayes, Decision Tree, Random Forest, and k-Nearest Neighbor) were assessed on a dataset of 206 songs. Music stimuli representing happy, sad, and fear as the predominant emotions. 78.44% is the highest accuracy value of from K-nearest neighbors yield. The study's findings are a promising first step toward using machine learning to analyse emotion in Sri Lankan traditional music.

The Fourier Transform was used to analyse and classify acoustic data gathered from defective air disc brakes in this article (FT). Two Norsonic Type 1228 microphones in a vehicle's laboratory were used to record the sound data. Matlab was used to examine the recorded data set that was transmitted to the computer via a data capture device.Mean, variance, entropy, and spectral rolloff of Fourier coefficients are Number of zero crossings that have been used as features in step to distinguish normal and faulty brakes. These features have been classified with 10x10 cross validation by using kth nearest neighbour algorithm with a success rate .

They provided an effective audio classification technique for detecting pulmonary edoema in this paper. The system uses a feature learning technique assumed on (NMF), then classified with logistic regression. Different NMF schemes were investigated and also compared with Principal Component Analysis compare. As a output, the proposed robust classification system based on NMF has been shown to be an effective method for audio-based pulmonary edoema identification. Tested using logistic regression and their effectiveness was studied Background noise collected from hospitals and speech from a speech corpus database was used to simulate loud data. This system using NMF on magnitude spectrum together with NMF signal enhancement proved to be a robust system under the outcome of environmental and speech noise, proving superior to other popular audio features, and feature selection techniques. If implemented in real-time, the proposed system can be useful in a number of ways like screening tool and in a hospital setting, it could aid physicians and nurses in screening patients might have excessive lung water. In a home environment, home-care practitioners and clients use the system to determine whether a trip to the doctor is warranted.

2.1 PROPOSED WORK

Our aim is to classify environmental audio sounds using deep learning techniques by MFCC features.

- Librosa and Scipy package are used to read the audio file.
- The sample rate of an audio file can be determined.
- To classify the audio sounds from metadata
- The feature are extract using Mel-Frequency Cepstral Coefficients from UrbanSound8K.

To rank and train the audio file using ANN model and to test the accuracy

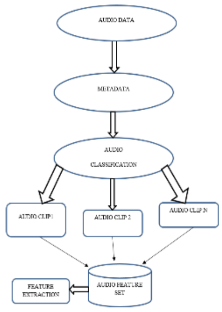

Fig -1: Block diagram of Audio data

The used dataset are bird sound dataset, Speech dataset, bird sound dataset in previous system. From the above listed dataset algorithm used in machine learning called ANN, NN, fourier Transformer and extracting features used as MFCC and GMM and these are coded in MATLAB. Then the trained model are driven in Matlab. Matlab coding is difficult, getting error while running the code. To overcome this, ANN algorithm is used in our project with the MFCC feature for extraction by using python programming.

Our project concepts applied to many practical scenarios e.g. to identify and group the various music clips using this models or classifying the voice from speakers, and use in robotics for auto analysis and performing the different task according to the user requirements.

Pandas is a popular python based upon data analysis tool kit which can be imported using import pandas as pd. t has a lot of features, such processing several file formats and converting a whole data table to NumPy.

This makes pandas a trusted ally in both data science and machine learning. Pandas hand out with data in one dimensional and two dimensional array. By calling the file name the output of first 10 files from metadata in urban sound 8K with the slice file name, free sound ID, starting and end of the frequency ,slice folder, class id, and class name of ten different sounds.

In pandas, series are referred to 1-D arrays

Dataframe are 2-D array

The numpy is a popular Python library. The numpy provides various types of array related operations in an easy-use way. Take steps to use numpy and it must be imported by using the "import numpy" statement.

The alias function in Python can be used to generate an alias, which is a shorter name for the supplied module.The statement can be used with the "as" statement in step to create alias as "np" for the numpy.

3. RESULT AND DISCUSSION

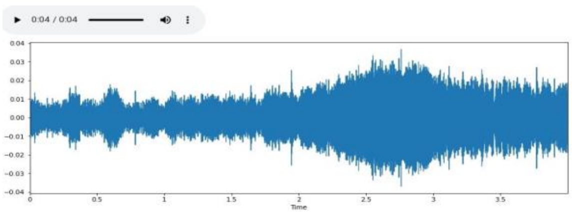

Fig -2: Representation of Children Playing

The above figure represents the audio file form folder 2-0Children playing.

- Filename:107653-9-012.wav
- Duration: 4 seconds
- Sample rate; 22050

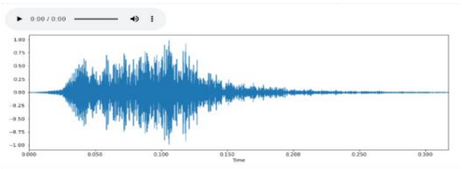

Fig -3: Representation of Dog Bark

Here, the loading of an audio file using the filename. The above figure represents the filename called Dog bark from Urbansound8K.There are 10 folders. Each sound is loaded using filename.

- Filename: 101415-3-0-8.wav
- Duration: 4 seconds
- Sample rate: 22050

AUDIO NAME	ACCURACY
Dog Barking	0.7538637518882751
Children Playing	0.7481396794319153

The Test accuracy for Dog Barking and Children Playing is 75% and 74% ,this accuracy vary for every iteration.

3. CONCLUSION

In our project, the audio file is read and loaded by using librosa and scipy package. The sample rate of an audio file is calculated in both the packages and also the classification of the audio sounds were done with the help of common space vector (metadata) file which was present inside from UrbanSound8K. The extracted features are trained using Artificial Neural Network (ANN) model and the process is repeated for various iterations. The accuracy result obtained was about 76%.

REFERENCES

1. K.Kumar and K.Chaturvedi, “An Audio Classification Approach using Feature Extraction Neural Network Classification Approach”, 2nd International Conference on Data, Engineering and Applications (IDEA),pp.16,2020.

2. M. M. M. Sukri, U. Fadlilah, S. Saon, A. K. Mahamad, M. M. Som and A. Sidek, "Bird Sound Identification based on Artificial Neural Network", IEEE Student Conference on Research and Development (SCOReD), pp. 342-345,2020.

.3. C. Joseph and S. Lekamge, "Machine Learning for Emotion Classification of Sri Lankan Folk Music”, 14th Conference on Industrial and Information Systems (ICIIS), pp. 360-365C,2019.

4. J. Nam, K. Choi, J. Lee, S. Chou and Y. Yang, "Deep Learning for Audio-Based Music Classification and Tagging: Teaching Computers to Distinguish Rock from Bach", Signal Processing Magazine, vol. 36, no. 1, pp. 41-51, Jan. 2019.

5. H. Purwins, B. Li, T. Virtanen, J. Schluter, S. Chang and T. Sainath, "Deep Learning for Audio Signal Processing”, in IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 2, pp. 206-219, May 2019.

6. Ertekln, Zeynep and Ozkurt, Nalan and Yilmaz, Cem ,”Analysis and Classification of Air disc Brake sounds in time and frequency domains”, Signal Processing and Communication Applications Conference (SIU), pp-1-4, 2018.

7. Kaur and R. Kumar, "Study and analysis of feature based automatic music genre classification using Gaussian mixture model”,International Conference on Inventive Computing and Informatics (ICICI), pp. 465-468,2017.

8. L Nanni, Y.M.G. Costa, D.R. Lucio, C.N. Silla Jr., S. Brahnam,” Combining visual and acoustic features for audio classification tasks”, Pattern Recognition Letters 88, pp. 49-56, 2017.

9. J.Sangeetha, R.Hariprasad, and S.Subhiksha, “Analysis of machine learning algorithms for audio event classification using Mel-frequency cepstral coefficients”, Applied Speech Processing, Primers in Biomediacl Imaging Devices and Systems,pp. 175-189, 2021.

10. F. Rong, "Audio Classification Method Based on Machine Learning”, International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), pp. 81-84, 2016.