International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

A Survey of Environmental Audio Classification

Nikitha N G[1], Nimisha M[2], Samantha Paul[3], Vibha Prasad[4], Radhika A D[5]

[1-4]Student, Department of Computer Science and Engineering, VVCE, Mysuru, Karnataka, India
[5]Assistant Professor, Department of Computer Science and Engineering, VVCE, Mysuru, Karnataka, India

---

Abstract - *In this literature survey paper, we have studied 12 papers regarding audio classification and their strength and weakness are also investigated. In this era of digital and electronic information, the audio data has been an significant part of many recent modern computer applications. The papers we have referred have an overview of existing techniques that are used to classify an audio clip. Here they have been systemized and analyzed their performances based on their theoretical considerations and run time performance. They have been experimented on different genres of varied categories thus illustrate the outcome of classification are significantly integral and effective.*

Key Words: **Audio classification, Deep neural network, Support vector machines, Feature extraction, GNN,K Nearest Neighbor(K-NN),Neural Network and naïve bayes(NB).**

1.INTRODUCTION

Audio Classification is the process of extracting and analyzing features from audio recordings.

A lot of considerable research is being carried out on the same topic due to the lack of work on audio event analysis. This topic is useful in a wide range of AI technologies such as Virtual Assistants, Text to Speech Applications and Automatic Speech Recognition Tasks.

We will analyze the proposed systems by several research papers and compare the outcomes. The previous researches are based on designing feature extraction systems to solve the challenging problems of audio classification.

The systems proposed previously learn useful features from data and act as a solid benchmark for future research.

2. LITERATURE SURVEY

[1] They have made use of Convolutional Deep Belief Networks that uses the probabilistic max-pooling as the building blocks for the network on unlabelled dataset for speech recognition and music recognition. This gave an accuracy of 72.7%. The limitation of this paper is its long training time.

[2] The use of Support Vector Machine (SVM) and Deep Neural Network (DNNhas been made to compare the classification by DNN, SVM and GMM. The combination of the DNN and the SVM classifiers provides a 6.5% improvement compared to DNN alone.

[3]The algorithm used for classification in this paper is Convolutional Neural Network (CNN). This was implemented on a labelled dataset. The fusion using different techniques works better than standalone classifiers and the accuracy ranges from 87.12% - 90.51% for different models.

[4] Use of Predictive Sparse Decomposition on unlabelled dataset has been implemented. The algorithm automatically learns necessary features from the provided audio data. The obtained accuracy on the GTZAN dataset is 92.7%.

[5]Convolutional neural networks are used in classifying short length audio clips of Environmental Sounds. A defined model of two complex layers is utilized The accuracy of CNN has been evaluated on three datasets of urban or environmental audio recordings. Here, the potential of convolutional neural network for Environmental Sound recognition and Classification is evaluated.

[6] An approach has been put forward for classification of environmental sound using 1D Convolution Neural Network and End to End Architecture, they have chosen this architecture to process audio signals of varied lengths, starting right from audio signal, a discriminative presentation gives a great classification efficiency on various natural sounds. The architecture of the network used consists of three to five convolutional layers, based on the length of the audio signal. This 1D Convolution Neural Network learns the filters from the audio waveform directly.

[7] Deep convolutional neural networks has been chosen to learn discriminative spectro-temporal patterns as they are very suitable to environmental sound classification in particular. However, a drawback of using this model is that it requires Labeled data which is scarce for this particular subject. This paper aims to overcome this problem by the using augmenting the audio data.

The utilization of CNN and data augmentation, the presented model yields top-notch results for environmental sound classification

[8] Using the historic sample data, the sample forgetting factor is fixed and the boundary vector is fetched. The model enforces the association between the KKT condition and the data to find the sample of the KKT condition which breach the decision function in the new data set. Final Decision function resulted is constantly rectified with the assistance of incremental learning until the end of the learning process to come up with the decision function, the system resulted in all in all average accuracy of 97.6%, with an error rate of 11.8. %.

[9] A procedure to categorize blended type of audio data, like speech combined with music with the help of a fuzzy rule based approach which inputs experimentally discovered thresholds, outputs a heuristic rule-based model for segregation into music, music, environmental sound, silence, background noise , etc. The model showcases that the fuzzy explanation of natural environment sound has a inverse impact on the efficiency of SVM and various classifiers. The experimental outcome proved that this system surpasses other classification systems in performance such as systems using k Nearest Neighbor (k-NN), Neural Network (NN), and Naive Bayes (NB).

[10]Many algorithms to categorize vocal clips into categories such as movies, sports, news, cartoons, etc are designed. They have used features like linear predictive coefficients, mel-frequency ceptral coefficients are fetched to categorize the audio clips at the end of the paper it shows that this method is effective and the accuracy rate is 92%.Finally from this paper we can conclude radial basis function neural network and super vector machines can be efficiently used for voice or audio classification.

[11] An attempt to highlight the importance of emotional sounds and its application in various media was made. The recognition of affect on musical sound is high compared to non-musical sounds. They felt there is still a lot of scope for work to enhance predictions made by neural network ,but were not clear of any solution. Even the output and methodology they used when they performed the research.

[12]It is an application of most of the artificial intelligence algorithms and their application. Its found that algorithms selected have impact system behavior, quality of human-robot interaction and computational transparency. Many learning models like deep learning, neural net work, hybrid and hidden markov models are reviewed and it shows that hybrid model have the best approach for sound processing.

3. CONCLUSIONS AND FUTURE WORK

In this literature survey, we have reviewed an evolutionary development in the field of audio recognition and classification. It has provided a discussion of audio features and classification techniques used in the various systems. To identify the limit of audio system, work of observation has been put on comparing the performance of automatic audio classification with human and manual performance as well as with other models of audio classification

From the evidences found on classification, it can be marked that there is still much room for further improvement over current automatic classification systems in terms of precision and accuracy.

REFERENCES

[1].Honglak Lee, Yan Largman, Peter Pham and Andrew Y .Ng"Unsupervised Learning of Sparse Features for Scalable Audio Classification"

[2]. ZviKonsand OrithToledo-Ronen"Audio Event Classification using Deep Neural Networks"

[3]. Loris Nania, GianlucaMagulo, Sheryl Brahnamb andMichelangelo Pacic"An Ensemble of Convolutional Neural Networks for Audio Classification"

[4].MikealHenaff, Kevin Jarett,KorayKavukcuogluand YannLeCun"Unsupervised Learning of Sparse Features for Scalable Audio Classification"

[5]. Karol J. Piczak, "environmental sound classification with convolutional neural networks"

[6]. SajjadAbdoli, Patrick Cardinal, Alessandro LameirasKoerich "end-to-end environmental sound classification using a 1d convolutional neural network"

[7]. Justin Salamon and Juan Pablo Bello "DEEP convolutional neural networks and data augmentation for environmental sound classification"

[8].Linyuan Fan "Audio example recognition and retrieval based on geometric incremental learning support vector machine"

[9]. Lei Chen, SuleGunduz and Tamer Ozsu "mixed type audio classification with support vector machine"

[10]. P. Dhanalakshmi, S. Palanivel, V.Ramalingam "Classification of audio signals using SVM and RBFNN"

[11]. Stuart Cunningham,Harrison Ridley, Jonathan Weinel, Richard Picking "Supervised machine learning for audio emotion recognition Enhancing film sound design

using audio features, regression models and artificial neural networks."

[12]. Troy Kelley, "A Review of Artificial Intelligence (AI) Algorithms for Sound Classification: Implications for Human–Robot Interaction".