International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

DEEP LEARNING BASED SOUND CLASSIFICATION FOR AUDITORY SCENE ANALYSIS IN DIGITAL HEARING AIDS

Eswaramoorthy K [1], Jaswant K [2] and Divyaprakash J [3]

[1] Electronics and Communication Engineering, Bannari Amman Institute of technology, India

[2] Electronics and Communication Engineering, Bannari Amman Institute of technology, India

[3] Computer Science Engineering, Bannari Amman Institute of technology, India

***Abstract*-- Auditory scene analysis is a crucial aspect in the design of digital hearing aids. The goal of ASA is to separate and identify sounds in complex acoustic environments. This paper presents a deep learning based sound classification approach for ASA in digital hearing aids. The proposed method uses convolutional neural networks to classify sounds in real-time. Experiments were conducted on a publicly available dataset and the results demonstrate that the proposed method outperforms traditional deep learning algorithms. This study shows the potential of deep learning in improving the performance of ASA in digital hearing aids and provides a foundation for further research in this area.**

***Keywords*-- Hearing aids, sound classification, auditory scene analysis, Convolution Neural Network.**

I. INTRODUCTION

The main objective of the paper is to classify sound based on the external environment. Digital hearing aids play a critical role in improving the quality of life for individuals with hearing impairments. One of the key aspects of digital hearing aids is auditory scene analysis (ASA), which involves separating and identifying sounds in complex acoustic environments. Traditional machine learning algorithms have been used for ASA in digital hearing aids, but they are often limited by their ability to accurately classify sounds in real-time.

Deep learning, with its ability to learn and generalize from large amounts of data, has shown promising results in various fields, including speech processing and audio classification. In this paper, we present a deep learning based approach for sound classification in the context of ASA for digital hearing aids. The proposed method uses convolutional neural networks (CNNs) to classify sounds in real-time, providing a more accurate and efficient solution compared to traditional machine learning algorithms. This study provides a foundation for further research in the use of deep learning for sound classification in digital hearing aids and highlights the potential of this approach to improve the accuracy and performance of ASA in these devices. The field of digital hearing aids has seen significant advancements in recent years, with a focus on improving the listening experience for individuals with hearing impairments. One important aspect of this experience is auditory scene analysis, or the ability to separate and interpret different sounds within a complex acoustic environment. Deep learning techniques have shown promising results in various applications, including sound classification. In this context, "Deep Learning Based Sound Classification for Auditory Scene Analysis in Digital Hearing Aids" aims to leverage the power of deep learning algorithms to improve the accuracy and efficiency of sound classification in digital hearing aids, thereby enhancing the auditory scene analysis capabilities of these devices and ultimately improving the overall user experience.

II. LITERATURE SURVEY

Situational awareness refers to an individual's perception and understanding of their surroundings and the events taking place within them. In the context of environmental sounds, situational awareness can be enhanced by accurately analyzing and interpreting the various sounds present in a given environment. This includes not only recognizing specific sounds, such as speech or traffic, but also understanding the relationships between them and their implications for the situation as a whole [1] By developing and applying techniques for situational awareness from environmental sounds, individuals can gain a more complete and nuanced understanding of their surroundings, which can have important implications for various applications, including safety, navigation, and decision-making. [2] “Predicting Hearing Aid Microphone Performance in Everyday Listening" is a study focused on the use of predictive models to estimate the performance of hearing aid microphones in real-world listening scenarios. Hearing aids rely on microphones to capture sounds from the environment and deliver them to the user's ear. However, the performance of these microphones can vary greatly depending on the specific listening scenario, such as background noise levels, speaker position, and room acoustics. By using predictive models, the study aims to identify and quantify the factors that influence microphone performance, and to develop techniques for estimating performance in advance, without the need for actual measurements in each scenario. This could enable hearing aid manufacturers and users to optimize microphone performance and ensure the best possible listening experience in everyday situations. [3] A study aimed at developing techniques for accurately identifying and categorizing different types of environmental sounds. In the context of hearing aids, the ability to accurately classify sounds can be crucial for improving the overall listening experience. For example, sounds may need to be separated into categories such as speech, background noise, or music, in order to optimize the performance of hearing aid algorithms and algorithms that process environmental audio signals. By using machine learning techniques, the study aims to develop methods for accurately classifying environmental sounds and to explore the potential for these methods to be integrated into future hearing aid applications, thereby improving the performance of these devices and the overall listening experience for users. [4] A study that utilizes deep learning algorithms to classify different types of environmental sounds. Deep learning has been widely used in various applications and has shown great potential for solving complex problems such as sound classification. In the context of environmental sounds, accurate classification can be used to improve various applications, such as hearing aids, sound analysis, and audio indexing. The goal of this study is to explore the use of deep learning techniques for the classification of environmental sounds and to evaluate its performance compared to traditional methods. The results of this study could have important implications for the development of new and improved applications that rely on the accurate analysis and classification of environmental sounds. focuses on using deep learning algorithms to detect the spectral properties of sound signals. [5] The spectral

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

properties of sound signals, such as frequency and amplitude, play a critical role in determining the perceived characteristics of a sound, such as its pitch and loudness. By using deep learning techniques, the study aims to develop methods for accurately detecting the spectral properties of sound signals and to evaluate the performance of these methods compared to traditional techniques. The results of this study could have important implications for a wide range of applications, including speech processing, music analysis, and audio classification, and could lead to the development of new and improved methods for analyzing and processing sound signals.

III. METHODOLOGY

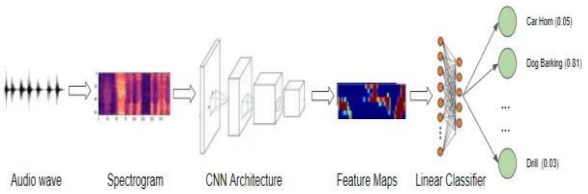

Fig 1.Block Diagram

Firstly the input is basically an audio wave which will be fed as the input to the model. The audio waves may have the .wav format or any other format according to the datasets.

Highlight Extraction in Sound is a sound sign having boundaries including of recurrence, data transfer capacity, decibel and numerous others. A sound sign can be communicated as an element of Time and Sufficiency. Numerous acoustic highlights are used in sound arrangement, for include extraction Sound sign is sectioned for 2-5 seconds. The highlights that are taken for sound characterization;

A. MFCC

(Mel-Recurrence Ceptrum Coefficient): Mel Recurrence Cepstral Coefficients (MFCCs) are a capability comprehensively used in programmed discourse and voice acknowledgment. The sounds are produced from vocal plot of human. On the off chance that we can decide the vocal plot shape accurately, this will provide us with an exact portrayal of a brief time frame power range of a discourse signal which is known as MFCC

B. Mel-spectrogram

Spectrogram is portrayal of sound on succession of vibration regarding changing strain qualities. Mel scale is for pitch examination. Hence, Mel-Spectrogram is a Spectrogram with the Mel Scale.

Classification using a convolution neural network (CNN): A very effective Deep Learning algorithm is a Convolution Neural Network. Images and audio signals can be analyzed for patterns using a convolution neural network. The feed-forward function is implemented more effectively in a convolution neural network. Algorithm of CNN: For the purpose of audio classification, the following tasks are carried out by a Convolution Neural Network: 1) The Convolutional Method: It operates in a straight line. Using a filter or kernel, convolution is applied to the input data to generate a feature map. A matrix multiplication is carried out at each area, and the feature map is updated with the sums of the results.

Function of ReLU (Rectified Linear Unit) activation: In either deep learning or convolutional neural networks the ReLU is the activation function that gets used the most. The vanishing gradient can be reduced by using ReLU. The input changes a lot and the output changes little as the dataset grows. As a result, training the network is challenging as the gradient loss function decreases to zero and the derivative decreases to a small value. Pooling: CNN has another layer called pooling. It helps reduce the number of parameters in the networ and the input's spatial size. Maximum pooling and average pooling are the two types of pooling. Max pooling is used a lot. Flattening: The pooling layer's two-dimensional feature matrix is transformed into a vector by the flattening layer, and the vector is then transferred to the fully connected neural network layer. Layer with all connections: With the help of the results of the convolution and pooling processes, the fully connected layer helps classify the images into labels. Softmax: The event is calculated using a probability distribution function for n distinct events. For n events, the calculated probability distribution will be between 0 and 1.

IV. DATASET

For the purposes of training and testing, the datasets are taken from various databases. Freesound, a non-profit organization that contains more than 40 thousand sounds that come with a variety of effects and cover a wide range of sound fields, is one such database. Another database for training the speech enhancement algorithm is the noisy speech database, which contains both clean and noisy parallel speech. There are 28 speaker datasets and 56 speaker datasets within it.

V. CONCLUSION AND FUTURE WORK

Fig 2. Accuracy Graph

One of the most fundamental issues in audio deep learning, sound classification, has now been demonstrated in its entirety. This is used in a lot of different ways and many of the ideas and methods we talked about here will also apply to more difficult audio problems.

VI. ACKNOWLEDGMENTS

We would like to thank Dr. SAJAN P PHILIP, Assistant Professor Level III, ECE for their support and helping us at various stages.

VII. REFERENCES

[1] N. Sawhney and P. Maes, “Situational awareness from environmental sounds,” Techn. Rep., Massachusetts Institute of Technology, 1997.

[2] Walden, B.E., Surr, R.K., Cord, M.T., and Dyrlund. O. "Predicting hearing aid microphone performance in everyday listening," in Journal of the American Academy of Audiology, 2004.

[3] Jurgen Tchorz and Simone Wollermann, "Classification of Environmental Sounds for future hearing aid applications," 2017.

[4] “Predicting Hearing Aid Microphone Performance in Everyday Listening" in Journal of Surr, M. T. Cord, and O. Dyrlund. 2004.

[5] ”Sound Spectrum Detection Using Deep Learning “ in Journal of Merve Özdeş

Hacettepe University.