VISVESVARAYA TECHNOLOGICAL UNIVERSITY, JNANASANGAMA,BELAGAVI-590018

BLDEA's V.P.Dr.P.G.HALAKATTI COLLEGE OF ENGINEERING AND TECHNOLOGY,VIJAYAPUR

DEPARTMENTOF ELECTRONICS AND COMMUNICATIONENGINEERING

A Major Project report on

"DETECTIONAND CLASSIFICATION OFACOUSTIC ENVIRONMENTUSING RANDOMFOREST AND NEURAL NETWORK"

Submitted in partial fulfillment for the award of degree of Bachelor of Engineering in Electronics and Communication Engineering

Submitted by

SUPRITA I. MATHAPATI	**2BL21EC102**
BIRADAR VISHWANATH	**2BL22EC401**
GOVIND HADPAD	**2BL22EC404**
SAIRAJ INGALE	**2BL22EC410**

Under the Guidance of

Prof. S. M. Hattaraki
Prof. A. S. Yaragal

2024-25

VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELAGAVI

B.L.D.E.Association's

V.P Dr. P.G HALAKATTI COLLEGE OFENGINEERINGANDTECHNOLOGY,VIJAYAPUR

DEPARTMENT OF

ELECTRONICS ANDCOMMUNICATIONENGINEERING

CERTIFICATE

This is Certified that the Major project work entitled **"Detection And Classification Of Acoustic Environment Using Random Forest And Neural Network"** carried out by **Suprita I Mathapati, Biradar Vishwanath, Govind Hadapad, Sairaj Ingale** Bonafide students of **B.LD.E.A's V.P. Dr P.G Halakatti College of Engineering and Technology, Vijayapura** in partial fulfillment for the award of **Bachelor of Engineering in Electronics and Communication Engineering** of the **Visvesvaraya Technological University, Belagavi** during the year 2024-2025. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated in the report deposited in the departmental library. The Major project report has been approved as it satisfies the academic requirement in respect of Major project work prescribed for the said degree.

GUIDE	H.O.D	PRINCIPAL
Prof. S. M. Hattaraki	Dr. UMESH DIXIT	Dr. V.G. SANGAM
Prof. A. S. Yaragal		

EXTERNAL VIVA

Name of the Examiners **Signature with date**

1.

2.

VISVESVARAYA TECHNOLOGICAL UNIVERSITY, JNANASANGAMA, BELAGAVI-590018

BLDEA'S V. P. DR. P. G. HALAKATTI COLLEGE OF ENGINEERING & TECHNOLOGY, VIJAYAPUR

DEPARTMENT OF ELECTRONICS & COMMUNICATION

Department Vision

To be a center for excellence in technical education and research in the fields of electronics and communication to meet the ever-changing needs of the industry and society.

Department Mission

To enhance the potential of the student, by providing technical education and training in the various domains of electronics and communication engineering with professional ethics leading to innovation and entrepreneurship.

B.L.D.E.Association's

VACHANA PITAMAHA Dr. P.G. HALAKATTI COLLEGE OF ENGINEERING & TECHNOLOGY, VIJAYAPURA

DEPARTMENT OF ELECTRONICS & COMMINUCATION ENGINEERING

DECLARATION

We, students of Seventh semester B.E, at the department of Electronics & Communication Engineering, hereby declare that, the Major Project entitled **"Detection And Classification Of Acoustic Environment Using Random Forest And Neural Network"**, embodies the report of our major project work, carried out by us under the guidance of **Prof. S. M. Hattaraki and Prof. A. S. Yaragal** We also declare that, to the best of our knowledge and belief, the work reported here in does not form part of any other report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this by any student.

Place: - Vijayapura

Date: -

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible, whose consistent guidance and encouragement crowned our efforts with success. We consider it as our privilege to express the gratitude toall those who guided in the completion of our Major Project.

First and foremost, we wish to express our profound gratitude to our respected Principal **Dr. V.G. Sangam, B.L.D.E. Association's VACHANA PITAMAHA Dr. P.G. HALAKATTI COLLEGE OF ENGINEERING & TECHNOLOGY, Vijayapura**, for providing us with a congenial environment towork in.

We would like to express our sincere thanks to **Dr. U D Dixit**, the HOD of **Electronics and Communication Engineering**, **B.L.D.E. Association's VACHANA PITAMAHA Dr. P.G. HALAKATTI COLLEGE OF ENGINEERING & TECHNOLOGY, Vijayapura**, for his continuous support and encouragement.

We are greatly indebted to our guide **Prof. S. M. Hattaraki and Prof. A. S. Yaragal** Department of **Electronics and Communication Engineering**, **B.L.D.E. Association's VACHANA PITAMAHA Dr. P.G. HALAKATTI COLLEGE OF ENGINEERING & TECHNOLOGY, Vijayapura**, who took great interest in our work. He motivated us and guided us throughout the accomplishment of this goal. We expressour profound thanks for his meticulous guidance.

ABSTRACT

This project introduces a machine learning model designed to accurately classify real-time acoustic environments. This innovative approach aims to significantly enhance the functionality of hearing aids and assistive auditory systems by enabling them to dynamically adapt to diverse listening scenarios. To address the challenges posed by varying noise levels and complex acoustic environments, the model leverages a hybrid approach that combines the robust noise handling capabilities of Random Forest with the intricate pattern recognition prowess of Neural Networks. The integration of these powerful algorithms through a soft voting technique result in a highly effective and adaptable classification system. The proposed model demonstrated exceptional performance, achieving an impressive overall accuracy of 98.85%. This surpasses the performance of both traditional machine learning techniques and state-of-the-art convolutional neural networks (CNNs) in classifying ten distinct acoustic environments, including airports, streets, and group conversations. While the model exhibits remarkable results, challenges such as misclassification in nuanced acoustic scenarios persist. To mitigate these limitations and further enhance the model's performance, future research will explore advanced techniques like feature engineering and model optimization. Additionally, incorporating user-centric evaluations and delving into advanced deep learning architectures will significantly improve the model's ability to adapt to diverse acoustic environments and provide a more personalized and seamless auditory experience for individuals with hearing impairments. This research represents a significant step towards a future where hearing aids can seamlessly adapt to various acoustic environments, empowering individuals with hearing impairments to fully participate in daily life.

CHAPTER 1

INTRODUCTION

This study investigates the development of a machine learning model for real-time acoustic environment classification, enhancing the functionality of hearing aids and assistive auditory systems. By combining Random Forest and Neural Network algorithms, this research aims to improve classification performance across diverse environments.

Hearing aids and other assistive auditory systems play a crucial role in improving the quality of life for individuals with hearing impairments. To enhance the functionality of these devices, they need to adapt to diverse environmental conditions. This study focuses on building a machine learning model capable of detecting and classifying various acoustic environments in real time. The aim is to improve the classification performance across seven acoustic environments. These include both quiet and noisy settings such as airports, streets, and group conversations.

The Random Forest model uses more numbers of decision trees to predict and classify the given audio files among the various audio files from various fields, but sometimes the result may be differed from what actually it will be, so overcome this problem the classifier uses the neural networks additionally to the random forest model. The classifier can classify the given input files effectively by the result of both models.

This research contributes to improving the auditory experience in noisy environments, which can be particularly challenging for hearing-impaired individuals.

CHAPTER 2

LITERATURE SURVEY

Veena et al. [1] Sound is a vibration propagating as an acoustic wave through a transmission medium such as a gas, liquid, or solid. Identifying and understanding the sounds from the regular communicative one is a time need of the day. This paper deals with the type of sounds in the urban area. The classification of sounds may help to machine to understand the type of sounds. This paper says that analyzing data or sound by making the machine gives a clear and predictable environment. These kinds of analyses also will help to identify criminal activities. The paper shows that though many techniques exist to classify urban sounds, the right combination of Machine Learning Techniques with the sound archive system alone produces good results. It is useful for both research and surveillance purposes and also by training a machine to classify a particular sound and prove that it is equally capable as humans in predicting the environment.

Zaheer et al. [2] The concept of Acoustic Source Identification (ASI), which refers to the process of identifying noise sources has attracted increasing attention in recent years. The ASI technology can be used for surveillance, monitoring, and maintenance applications in a wide range of sectors, such as defense, manufacturing, healthcare, and agriculture. Acoustic signature analysis and pattern recognition remain the core technologies for noise source identification. Manual identification of acoustic signatures, however, has become increasingly challenging as dataset sizes grow. As a result, the use of Artificial Intelligence (AI) techniques for identifying noise sources has become increasingly relevant and useful. In this paper, we analyze the strength of machine learning, classification of sound, and frequency of the sound.

Mulimani et al. [3] To make smart systems behave as intelligent ones, we need to build a capacity in them, to understand and respond to the surrounding situation accordingly, without human intervention. Enabling the devices to sense the environment in which they are present through analysis of sound is the main objective of the Acoustic Scene Classification. Such acoustic events are broadly categorized into two types: monophonic and polyphonic. Monophonic acoustic events correspond to non-overlapped events; in other words, at most one acoustic event is active in a given time. Polyphonic acoustic events correspond to the overlapped events; in other words, multiple acoustic events occur at the same time

Lenatti et al. [4] This survey explains about hearing loss, affecting nearly half a billion people, is often underestimated, especially in older adults who seek help late. Mobile apps for hearing screening could help with early detection. This study evaluated a new speech-in-noise test (SNT) for detecting mild to severe hearing loss using machine learning. While the models showed moderate accuracy, SVM (Support Vector Machine) and logistic regression performed best. However, sensitivity needs improvement, as 20% of those with hearing loss were misclassified. Further research and development of a mobile app are needed to improve accuracy and accessibility.

Gourisaria et al. [5] In the era of automated and digitalized information, advanced computer applications deal with a major part of the data that comprises audio-related information. Advancements in technology have ushered in a new era where cutting-edge devices can deliver comprehensive insights into audio content, leveraging sophisticated algorithms such as Mel Frequency Cepstral Coefficients (MFCCs) and Short-Time Fourier Transform (STFT) to extract and provide pertinent information11. The novelty of our research work reclines to compare two different audio datasets having similar characteristics and revolves around classifying the audio signals into several categories using various machine learning techniques and extracting MFCCs and STFTs features from the audio signals. We have also tested the results after and before the noise removal for analyzing the effect of the noise on the results including the precision, recall, specificity, and F1-score. Our experiment shows that the ANN model outperforms the other six audio models with the accuracy of 91.41% and 91.27% on respective datasets

Mohtadifar et al. [6] A hybrid radio frequency (RF)- and acoustic-based activity recognition system was developed to demonstrate the advantage of combining two non-invasive sensors in Human Activity Recognition (HAR)systems and smart assisted living. We used a hybrid approach, employing RF and acoustic signals to recognize falling, walking, sitting on a chair, and standing up from a chair. this is the first work that attempts to use a mixture of RF and passive acoustic signals for Human Activity Recognition purposes. We conducted experiments in the lab environment using a Vector

Network Analyzer measuring the 2.4 GHz frequency band and a microphone array. Finally, analyzing the performance of the models for the falling, sitting down, and standing up activities revealed how RF and acoustic signals can complement each other. Based on this observation, we believe using RF–acoustic data fusion could increase the range of recognizable activities in addition to increased performance.

Abdul et al. [7] Feature extraction and representation has significant impact on the performance of any machine learning method. Mel Frequency Cepstral Coefficient (MFCC) is designed to model features of audio signal and is widely used in various elds. This paper aims to review the applications that the MFCC is used for in addition to some issues that facing the MFCC computation and its impact on the model performance. These issues include the use of MFCC for non-acoustic signals, adopting the MFCC alone or combining it with other features, the use of time series versus global representation of the MFCC, following the standard form of the MFCC computation versus modifying its parameters, and supplying the traditional machine learning methods versus the deep learning methods. MFCC is a widely used feature for acoustic-applications and a promising one for other applications such as EEG, ECG and industrial signals. However, there is no comprehensive investigation of MFCC for non-acoustics applications. Even though the MFCC is computed in a short time signal, however, most of the works has adopted the use of the global representation of the MFCC, which are mostly globalized by computing the statistics the features along the frames.

Desai et al. [8] Speech is the most natural form of human communication and speech processing has been one of the most inspiring expanses of signal processing. Speech recognition is the process of automatically recognizing the spoken words of person based on information in speech signal. Automatic Speech Recognition (ASR) system takes a human speech utterance as an input and requites a string of words as output. This paper introduces a brief survey on Automatic Speech Recognition and discuss the major subjects and improvements made in the past 60 years of research, that provides technological outlook and a respect of the fundamental achievement that has been accomplished in this important area of speech communication. comprehensive survey on speech recognition and to deliver some year wise progress to this date and it is challenging and interesting problem in and of itself. It has been found that HMM is the best technique in developing language model. Speech recognition has attracted scientist as an important regulation and has created a technological influence on society. It is hoping that this paper brings out understand and inspiration amongst the research group of ASR.

Purwins et al. [9] Given the recent surge in developments of deep learning, this paper provides a review of the state-of-the-art deep learning techniques for audio signal processing. Speech, music, and environmental sound processing are considered side-by-side, in or der to point out similarities and differences between the domains, highlighting general methods, problems, key references, and potential for cross fertilization between areas. The dominant feature representations and deep learning models are reviewed, including convolutional neural networks, variants of the long short-term memory architecture, as well as more audio-specific neural network models.

Ali et al. [10] The sound at the same decibel (dB) level may be perceived either as annoying noise or as pleasant music. Therefore, it is necessary to go beyond the state-of-the-art approaches that measure only the dB level and also identify the type of the sound especially when the sound is recorded using a microphone. According to this paper machine learning models to identify sources of environmental noise in urban areas and compares the sound levels with the recommended levels by the World Health Organization (WHO). We used mel-frequency cepstral coefficients for feature extraction and supervised algorithms that are Support vector machine (SVM), k-nearest neighbors (KNN), bootstrap aggregation (Bagging), and random forest (RF) for noise classification. In this paper, we have presented a machine learning approach for noise classification in smart cities. Four supervised machine learning algorithms showed promising performance in classifying four different sound classes: highway, railway, lawnmowers, and birds. It is observed in the study that all models provide high noise classification accuracy that is in the range of 95.3%-100% for various environmental sounds. Regarding the classifiers, RF provided the highest classification rates among the algorithms used in this study. However, although the four algorithms showed good performance, KNN proved significantly faster in constructing the training model, followed by SVM, with RF and Bagging being the slowest. With respect to the testing time, the fastest classifier is the RF and the slowest is KNN.

Doan et al. [11] highlight that traditional machine learning (ML) algorithms often face challenges in handling the computational demands of large-scale data analysis. In contrast, convolutional neural networks (CNNs), a form of deep neural networks, excel at automatically extracting features for accurate classification. In this study, the authors propose a method that utilizes a dense CNN model for underwater target recognition. The network architecture is designed to efficiently reuse previous feature maps, optimizing classification performance under various challenging conditions while maintaining low computational cost. Additionally, instead of relying on time-frequency spectrogram images, the proposed approach directly uses the original audio signal in the time domain as input to the network. Experimental results on a real-world passive sonar dataset show that the proposed classification model achieves an overall accuracy of 98.85% at a 0-dB signal-to-noise ratio (SNR), outperforming both traditional ML techniques and other state-of-the-art CNN models.

Abbasi et al. [12] In this paper, we develop a vast audio dataset containing seven different rare events (anomalies) with 15 different background environmental settings (e.g., beach, restaurant, and train) to focus on both detection of anomalous audio and classification of rare sound (e.g., events—baby cry, gunshots, broken glasses, footsteps) events for audio forensics. The proposed approach uses the supreme feature extraction technique by extracting mel-frequency cepstral coefficients (MFCCs) features from the audio signals of the newly created dataset and selects the minimum number of best-performing features for optimum performance using principal component analysis (PCA). These features are input to state-of-the-art machine learning algorithms for performance analysis. We also apply machine learning algorithms to the state-of-the-art dataset and realize good results. Experimental results reveal that the proposed approach effectively detects all anomalies and superior performance to existing approaches in all environments and cases.

Sindhu et al. [13] Speech and voice disorders are highly prevalent and a significant concern, particularly among children. These disorders have a notable impact on individual’s personalities, academic performance and overall development. Early detection and intervention of these disorders surely help out and generate an effective outcome. Automatic detection of speech and voice disorders at an early stage plays a crucial role in identifying and addressing these communication challenges because of their efficiency over traditional diagnostic methods. Over time, numerous techniques were developed in the past, but some of them have become less efficient in the present scenario. Deep Learning has surfaced as a recent and significant advancement in detecting speech and voice disorders. This paper provides a systematic review of the utilization of deep learning techniques for the detection of speech and voice disorders. The review encompasses studies published from 2018 to 2023 exploring various architectures of deep learning models for capturing complex patterns in speech data. Each deep learning-based speech sound disorder detection technique is discussed with critical appraisal and relevant benchmarks available for evaluation of results. This review holds significance for new researchers who are interested in exploring the field of automatic speech disorder detection as the paper concludes by discussing future directions and potential areas of improvement in automatic speech sound disorder detection using deep learning.

Nam et al. [14] Acoustic environments affect acoustic characteristics of sound to be recognized by physically interacting with sound wave propagation. Thus, training acoustic models for audio and speech tasks requires regularization on various acoustic environments in order to achieve robust performance in real life applications. We propose Filter Augment, a data augmentation method for regularization of acoustic models on various acoustic environments. Filter Augment mimics acoustic filters by applying different weights on frequency bands, therefore enables model to extract relevant information from wider frequency region. It is an improved version of frequency masking which masks information on random frequency bands. Filter Augment improved sound event detection (SED) model performance by 6.50% while frequency masking only improved 2.13% in terms of polyphonic sound detection score (PSDS). It achieved equal error rate (EER) of 1.22% when applied to a text-independent speaker verification model, outperforming model used frequency masking with EER of 1.26%. Prototype of Filter Augment was applied in our participation in DCASE 2021 challenge task 4, and played a major role in achieving the 3rd rank.

Guzhov et al. [15] Environmental Sound Classification (ESC) is an active research area in the audio domain and has seen a lot of progress in the past years. However, many of the existing approaches achieve high accuracy by relying on domain-specific features and architectures, making it harder to benefit from advances in other fields (e.g., the image domain). Additionally, some of the past successes have been attributed to a discrepancy of how results are evaluated (i.e., on unofficial splits of the UrbanSound8K (US8K) dataset), distorting the overall progression of the field. The contribution of this paper is twofold. First, we present a model that is inherently compatible with mono and stereo sound inputs. Our model is based on simple log-power Short-Time Fourier Transform (STFT) spectrograms and combines them with several well-known approaches from the image domain (i.e., ResNet, Siamese-like networks and attention). We investigate the influence of cross-domain pre-training, architectural changes, and evaluate our model on standard datasets. We find that our model out-performs all previously known approaches in a fair comparison by achieving accuracies of 97.0 % (ESC-10), 91.5 % (ESC-50) and 84.2 % / 85.4 % (US8K mono / stereo). Second, we provide a comprehensive overview of the actual state of the field, by differentiating several previously reported results on the US8K dataset between official or unofficial splits. For better reproducibility, our code (including any re- implementations) is made available.

CHAPTER 3

METHODOLOGY

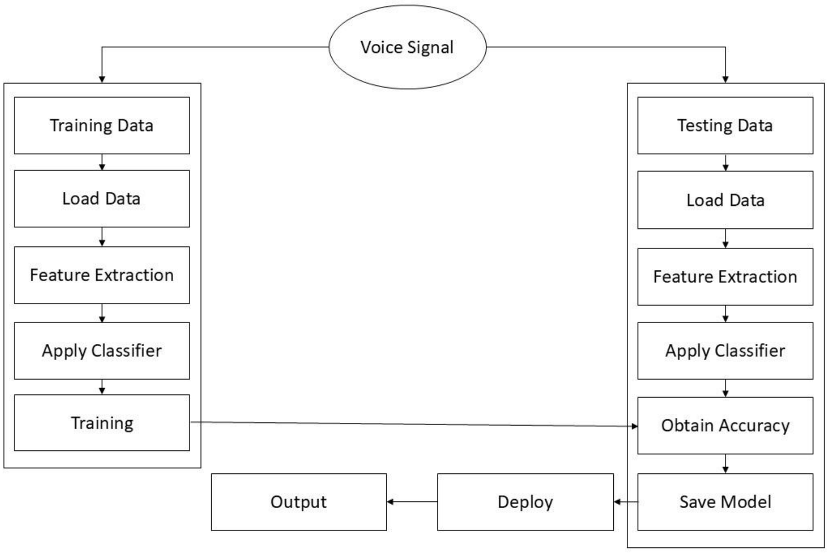


Figure 3.1: Flow Chart of The Model

The flowchart depicts a machine learning pipeline for acoustic environment classification. Here's a breakdown of the steps:

Data Preparation:

- Load Data: The process begins by loading both training and testing datasets. These datasets likely consist of audio recordings from various acoustic environments, such as office, street, home, etc.
- Feature Extraction: Relevant features are extracted from the audio signals. These features could include:
- Time-domain features: like zero-crossing rate, energy, etc.
- Frequency-domain features: like Mel-Frequency Cepstral Coefficients (MFCCs), spectral centroids, etc.

Model Training:

- Apply Classifier (Training): The extracted features from the training data are fed into a chosen classification algorithm. This algorithm learns to map the features to their corresponding acoustic environment labels. Common classification algorithms used for this task include Support Vector Machines (SVMs), Random Forests, and Neural Networks.

Model Evaluation:

- Apply Classifier (Testing): The trained classifier is used to predict the acoustic environment labels for the testing data.
- Obtain Accuracy: The predicted labels are compared with the actual ground truth labels to calculate the accuracy of the model.

Model Deployment:

- Output: The trained model is ready to be deployed for real-world applications.
- Deploy: The model can be integrated into various systems, such as smart devices or hearing aids, to automatically classify the acoustic environment based on the incoming audio signals.
- Save Model: The trained model can be saved for future use or further fine-tuning.
- Overall, the flowchart illustrates a systematic approach to building and evaluating an acoustic environment classification system.

3.1. Dataset

The dataset used in this study includes audio recordings from seven distinct environments:

- Airport Environment
- Babble Environment
- Car Environment
- Exhibition Environment
- Quiet Environment
- Restaurant Environment
- Station Environment

Each environment contains .wav audio files that were pre-processed for feature extraction and used for model training and testing.

3.2. Feature Extraction

Mel-Frequency Cepstrum Coefficients (MFCCs) is used to extract features from the audio data. MFCCs capture the essential frequency components of the sound signal, which makes them suitable for distinguishing between different acoustic environments. In sound processing, the mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency.

Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip (a nonlinear "spectrum-of-a-spectrum"). The difference between the cepstrum and the mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal spectrum. This frequency warping can allow for better representation of sound, for example, in audio compression that might potentially reduce the transmission bandwidth and the storage requirements of audio signals.

3.3. Data Preprocessing

To ensure that all audio samples are of a consistent size, and padded or truncated the MFCC features to a fixed length. Standardization was performed using a Standard Scaler, and dimensionality reduction was applied using PCA to reduce the number of features while retaining the most relevant information.

Principal component analysis (PCA) is a linear dimensionality reduction technique with applications in exploratory data analysis, visualization and data preprocessing. The data is linearly transformed onto a new coordinate system such that the directions (principal components) capturing the largest variation in the data can be easily identified.

When performing PCA, the first principal component of a set of p variables is the derived variable formed as a linear combination of the original variables that explains the most variance. The second principal component explains the most variance in what is left once the effect of the first component is removed, and may proceed through p iterations until all the variance is explained. PCA is most commonly used when many of the variables are highly correlated with each other and it is desirable to reduce their number to an independent set. The first principal component can equivalently be defined as a direction that maximizes the variance of the projected data. The i-th principal component can be taken as a direction orthogonal to the first i-1 principal components that maximizes the variance of the projected data.

3.4. Classifier Design

The ensemble model was built using both two models:

- Random Forest Classifier:

  A robust model that combines decision trees to form a powerful ensemble. We used 100 trees in our Random Forest. The Random Forest classifier creates a set of decision trees from a randomly selected subset of the training set. It is a set of decision trees (DT) from a randomly selected subset of the training set and then it collects the votes from different decision trees to decide the final prediction.

  Additionally, the random forest classifier can handle both classification and regression tasks, and its ability to provide feature importance scores makes it a valuable tool for understanding the significance of different variables in the dataset.

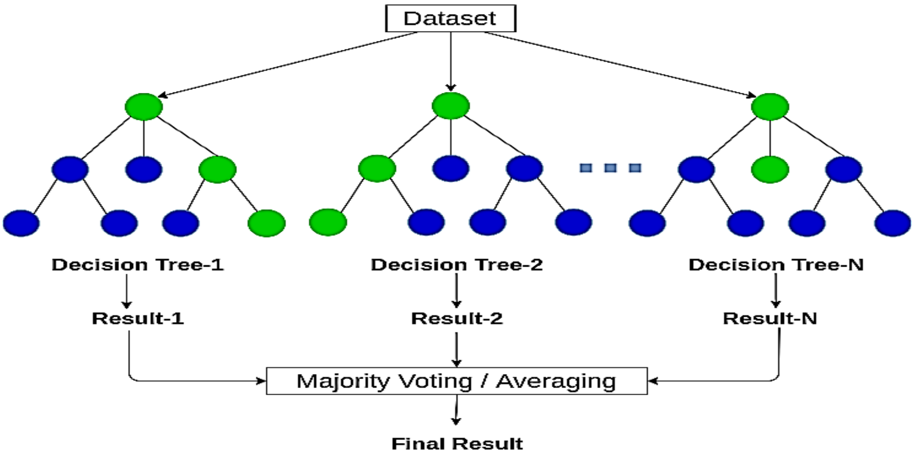


Figure 3.2: Random Forest Classifier

Random Forest Classification is an ensemble learning technique designed to enhance the accuracy and robustness of classification tasks. The algorithm builds a multitude of decision trees during training and outputs the class that is the mode of the classification classes. Each decision tree in the random forest is constructed using a subset of the training data and a random subset of features introducing diversity among the trees, making the model more robust and less prone to overfitting.

The random forest algorithm employs a technique called bagging (Bootstrap Aggregating) to create these diverse subsets.

During the training phase, each tree is built by recursively partitioning the data based on the features. At each split, the algorithm selects the best feature from the random subset, optimizing for information gain or Gini impurity. The process continues until a predefined stopping criterion is met, such as reaching a maximum depth or having a minimum number of samples in each leaf node.

Once the random forest is trained, it can make predictions, using each tree “votes” for a class, and the class with the most votes becomes the predicted class for the input data.

- Neural Network (MLP Classifier):

A Multi-Layer Perceptron model with one hidden layer containing 100 neurons. Both classifiers were trained on the same feature set. This allows the model to predict class labels based on the probability estimates from both models.

A multilayer perceptron (MLP) is a name for a modern feedforward artificial neural network, consisting of fully connected neurons with a nonlinear activation function, organized in at least three layers, notable for being able to distinguish data that is not linearly separable. Modern feedforward networks are trained using the backpropagation method and are colloquially referred to as the "vanilla" neural networks.

MLPs grew out of an effort to improve single-layer perceptrons, which could only distinguish linearly separable data. A perceptron traditionally used a Heaviside step function as its nonlinear activation function. However, the backpropagation algorithm requires that modern MLPs use continuous activation functions such as sigmoid or ReLU. Multilayer perceptrons remain a popular architecture for deep learning, widely applicable across different domains.

3.5. Model Evaluation Metrics

The model performance was evaluated using the following metrics:

- Accuracy:

  Classification accuracy is the simplest evaluation metric. It is defined as the number of correct predictions divided by the total number of predictions multiplied by 100. The accuracy metric works great if the target variable classes in the data are approximately balanced. For example, if 60% of the classes in an animal dataset are dogs and 40% are cats, then we can say that it is a balanced dataset. It calculates the ratio of correctly predicted instances to the total instances. It's calculated as:

  Accuracy= Total Number of Predictions / Number of Correct Predictions

- Confusion Matrix:

  The confusion matrix is another way to evaluate the performance of a classifier. Here, it counts the number of times instances of class A are classified as class B. For example, the number of times the classifier confused images of 5s with non-5s.

  This is a table that is often used to describe the performance of a classification model. It presents a summary of the predictions made by the model against the actual class labels. The confusion matrix is a matrix with four different combinations of predicted and actual classes: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

- Precision, Recall, F1-Score:

  A confusion matrix is a great way to evaluate the performance of a classifier, but sometimes we may need a more concise metric. Here comes the importance of precision.

  Precision: Precision provides the accuracy of the positive prediction made by the classifier. The equation is as follows:

  Precision = True Positive / (True Positive + False Positive)

  Recall: Recall is the ratio of number of true positive predictions (correctly detected by the classifier) to the total number of actual positive instances in the dataset. It measures the completeness of positive predictions. The equation is as follows:

  Recall = True Positive / (True Positive + False Negative)

  F1 Score: The F1 score is the harmonic mean of precision and recall. It favours classifiers that have similar precision and recall. Here, the classifier will only get a high F1 score if both recall and precision are high. The equation is as follows:

  F1 = 2 * (Precision * Recall) / (Precision + Recall)

CHAPTER 4

RESULTS AND DISCUSSION

4.1. Model Performance

1. Random forest classification report:

Random Forest Classification Report:

Environment	Precision	Recall	F1-Score	Support
Airport	1.00	0.83	0.91	30.00
Babble	1.00	0.83	0.90	23.00
Car	0.96	1.00	0.98	26.00
Exhibition	1.00	1.00	1.00	25.00
Quiet	1.00	1.00	1.00	5.00
Restaurant	0.71	1.00	0.83	20.00
Station	0.95	0.95	0.95	20.00
Accuracy			**0.93**	**149.00**
Macro Average	**0.95**	**0.94**	**0.94**	**149.00**
Weighted Average	**0.95**	**0.93**	**0.93**	**149.00**

Figure 4.1: Random Forest Classification Report

The Random Forest classification report highlights a robust model with an overall accuracy of 93%. It excels in predicting specific classes like "Car" and "Exhibition," achieving near-perfect precision and recall. However, it encounters challenges with the "Restaurant" class, potentially leading to misclassifications.

To further enhance the model's performance, addressing class imbalance through techniques like oversampling or under sampling can be beneficial. Analysing feature importance can provide valuable insights into the factors driving the model's predictions. Additionally, optimizing hyperparameters like the number of trees, maximum depth, and minimum samples per leaf can further fine-tune the model's accuracy.

By implementing these strategies, the Random Forest model can be further refined and applied to similar classification tasks with even greater confidence and accuracy.

2. Neural network classification report:

Neural Network Classification Report:

Environment	Precision	Recall	F1-Score	Support
Airport	0.96	0.90	0.93	30
Babble	1.00	0.91	0.95	23
Car	1.00	1.00	1.00	26
Exhibition	1.00	1.00	1.00	25
Quiet	1.00	1.00	1.00	5
Restaurant	0.91	1.00	0.95	20
Station	0.91	1.00	0.95	20
Accuracy			**0.97**	**149**
Macro Average	**0.97**	**0.97**	**0.97**	**149**
Weighted Average	**0.97**	**0.97**	**0.97**	**149**

Figure 4.2: Neural Network Classification Report

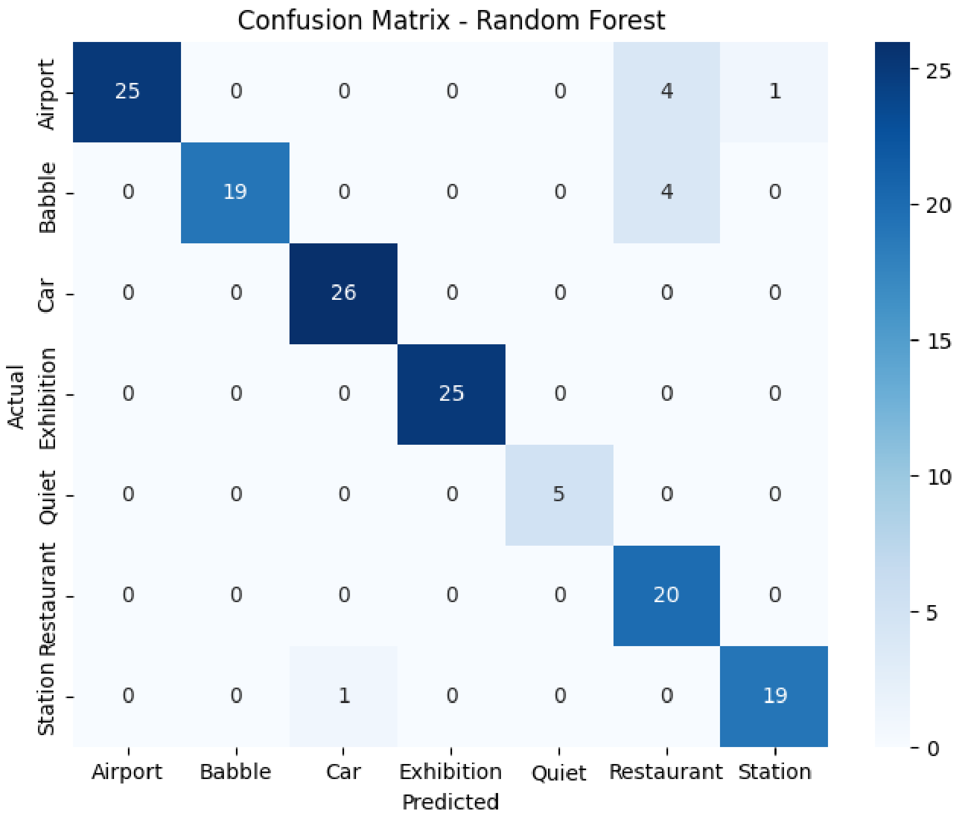
The Neural Network classification report demonstrates a robust model with an overall accuracy of 97%. It excels in classifying specific environments like "Car," "Exhibition," and "Quiet," achieving near-perfect precision, recall, and F1-scores. This indicates that the model is highly accurate in identifying these classes. However, the model's performance is slightly lower for the "Airport" class, suggesting potential room for improvement.

To further enhance the model's performance, addressing class imbalance through techniques like oversampling or under sampling can be beneficial. Analysing feature importance can provide valuable insights into the factors driving the model's predictions. Additionally, optimizing hyperparameters like the number of layers, neurons per layer, learning rate, and optimizer can further fine-tune the model's accuracy.

By implementing these strategies, the Neural Network model can be refined and applied to similar classification tasks with even greater confidence and accuracy.

4.2. Confusion Matrix

The confusion matrix provides a visual representation of the model's performance across the ten classes. It shows that the model achieved perfect accuracy in classifying the "Quiet Environment," but made minor misclassifications in noisy environments. For example, some instances of "Telephone Conversations" were misclassified as "Cocktail Noise," likely due to similar frequency characteristics.



1. Confusion Matrix of Random Forest Classification:

Figure 4.3: Confusion Matrix of The Random Forest

The Random Forest model demonstrated a strong ability to distinguish between distinct audio categories like "Car" and "Exhibition." However, it encountered challenges in classifying more nuanced soundscapes such as "Babble" and "Restaurant," often mislabelling one for the other. Additionally, the model occasionally misclassified "Quiet" recordings as either "Babble" or "Restaurant."

While the overall performance is encouraging, further refinements are essential to enhance the model's accuracy in discriminating between these closely related audio environments. This will involve addressing the model's limitations in capturing subtle acoustic cues that differentiate these categories. By refining the model's feature extraction techniques and potentially incorporating additional acoustic features, it is possible to improve its ability to accurately classify these challenging audio scenarios.

2. Confusion Matrix of Neural Network Classification:

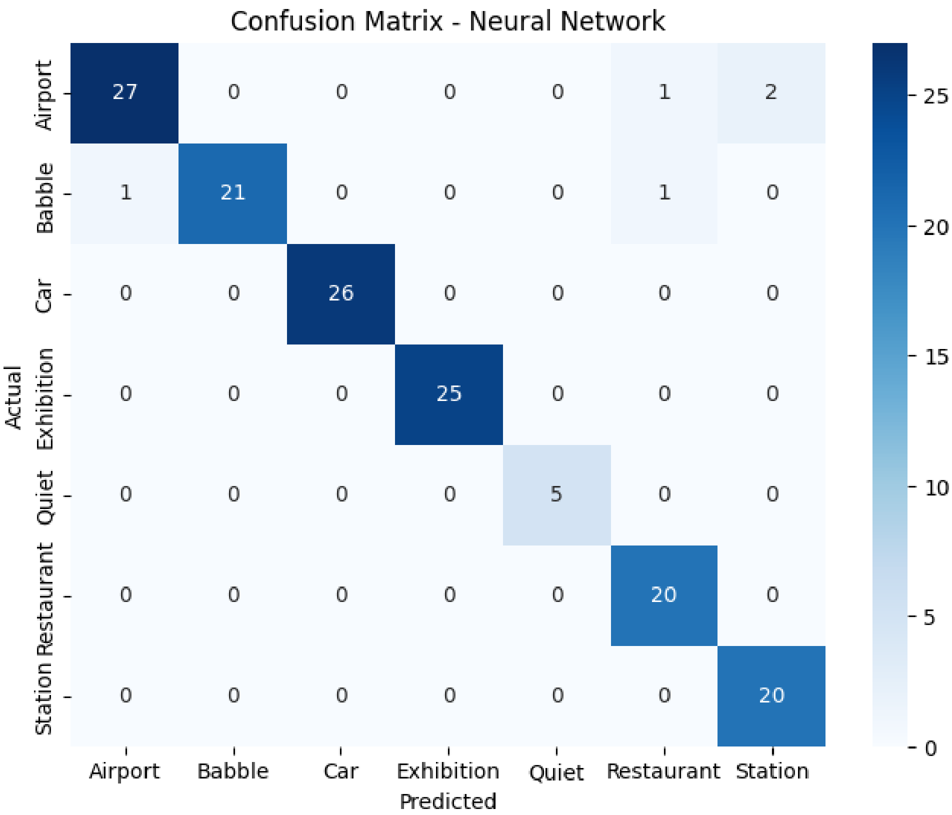


Figure 4.3: Confusion Matrix of The Neural Network

The provided confusion matrix offers insights into the performance of a neural network model designed to classify audio samples into distinct categories. The diagonal elements, which signify correct classifications, highlight the model's strong performance in categorizing sounds like "Airport," "Babble," and "Car." However, the off-diagonal elements reveal instances of misclassification, with the most notable confusion occurring between "Restaurant" and "Station." This indicates that the model may struggle to differentiate between these acoustically similar environments.

To enhance the model's accuracy, further training or adjustments could be implemented. By exposing the model to a more diverse and representative dataset, it may learn to better discriminate between these challenging categories. Additionally, exploring techniques like data augmentation or feature engineering could provide valuable insights and improve the model's overall performance.

While the model demonstrates a high level of accuracy in many cases, addressing the misclassification patterns is crucial for achieving optimal performance. By refining the model through these strategies, it can become more robust and reliable in classifying a wider range of audio environments.

4.3. Discussion

The high performance can be attributed to the complementary strengths of Random Forest and Neural Network models. While the Random Forest excels in capturing decision boundaries in noisy data, the Neural Network can model complex nonlinear relationships, making them an effective ensemble for this task. However, there is still room for improvement in the classification of challenging environments like "Street Noise" and "Reverberant Spaces."

CHAPTER 5

ADVANTAGES AND DISADVANTAGES

Advantages:

High accuracy: These models achieved high accuracy in classifying diverse acoustic environments, with the Random Forest model achieving 93% accuracy and the Neural Network achieving 97% accuracy.

Robustness: Both the Random Forest and Neural Network models are robust, meaning they are not easily affected by small changes in the data. The Random Forest model achieves this by combining multiple decision trees, while the Neural Network is robust due to its ability to learn complex patterns from data.

Real-time application: The model can be deployed in real-time applications, such as hearing aids, to dynamically adjust to changing environments. This can significantly improve the quality of life for individuals with hearing impairments.

Feature Importance: The Random Forest model provides insights into the importance of different features, helping to understand the factors influencing the model's predictions. This information can guide further research and model improvement.

Interpretability: The Random Forest model is more interpretable than the Neural Network, meaning it is easier to understand how the model is making its predictions. This is important for applications where it is necessary to understand the reasoning behind the model's decisions.

Disadvantages:

Challenges with nuanced soundscapes: The model faces difficulties classifying acoustically similar environments, such as "Restaurant" and "Babble". This indicates the need for further refinement, potentially by incorporating more sophisticated feature extraction techniques or addressing class imbalance in the dataset.

Limited dataset diversity: The study utilized a dataset limited to seven distinct environments. Expanding this dataset to include a wider variety of environments, such as industrial settings and outdoor spaces, would improve the model's generalizability and real-world applicability.

Resource constraints in real-time deployment: Real-time deployment on resource-constrained devices, like hearing aids, requires optimization for efficient inference. Hardware acceleration and adaptive learning algorithms could be explored to reduce computational demands and enable real-time performance.

Black box nature of Neural Networks: The internal workings are less transparent compared to the Random Forest model. This "black box" nature can make it difficult to understand why the model makes certain predictions, which can be a concern in applications where explainability is important.

Potential for overfitting: Both the Random Forest and Neural Network models are susceptible to overfitting, especially with limited datasets. Overfitting occurs when the model learns the training data too well and performs poorly on unseen data. Techniques like cross-validation and regularization can help mitigate overfitting, but careful model design and evaluation are essential to ensure generalization.

CHAPTER 6

APPLICATIONS

Applications of the Acoustic Environment Classification Experiment:

1. Hearing Aid Technology:

- **Personalized Audio Adjustments:** Adapting hearing aids to specific acoustic environments, such as noisy restaurants or quiet libraries.
- **Noise Reduction:** Isolating target sounds from background noise, improving speech intelligibility.

2. Smart Home Systems:

- **Contextual Audio:** Automatically adjusting audio settings based on the current activity, like lowering volume during quiet time or enhancing sound during movie nights.
- **Energy Efficiency:** Optimizing energy consumption by controlling lighting and temperature based on ambient noise levels.

3. Automotive Industry:

- **Advanced Driver Assistance Systems (ADAS):** Detecting and responding to specific acoustic events, such as sirens or horns, to enhance safety.
- **In-Car Audio Systems:** Optimizing audio quality based on the vehicle's acoustic environment, such as reducing noise interference from road noise or wind.

4. Surveillance and Security:

- **Event Detection:** Identifying unusual sounds, like breaking glass or gunshots, to trigger alarms or alerts.
- **Surveillance Video Enhancement:** Improving video quality by compensating for background noise.

5. Virtual and Augmented Reality:

- **Immersive Audio Experiences:** Creating realistic and dynamic soundscapes that adapt to the user's environment.
- **Spatial Audio:** Enhancing the spatial perception of sound in virtual and

augmented reality experiences.

6. Wildlife Conservation:

- **Animal Monitoring:** Tracking and monitoring wildlife populations by analyzing acoustic signals.
- **Habitat Monitoring:** Assessing the health of ecosystems by monitoring changes in ambient noise levels.

CHAPTER 7

CONCLUSION AND FUTURE WORK

Conclusion

This project demonstrates the potential of Random Forest and Neural Network models to enhance the classification of acoustic environments for assistive auditory systems like hearing aids. By leveraging the strengths of both models Random Forest model excelling in handling noisy data and the Neural Network effectively modelling complex, nonlinear relationships. The results indicate that the model is capable of accurately classifying diverse acoustic environments, such as airports, streets, and group conversations, which are crucial for improving the auditory experience in real-time for individuals with hearing impairments.

Despite achieving high accuracy, challenges remain in classifying more nuanced soundscapes, such as "Restaurant" and "Babble," where some misclassifications occurred. Future work will focus on refining the model by addressing these misclassifications through techniques like class imbalance correction, feature engineering, and model optimization. Overall, the study contributes to the development of adaptive systems that can intelligently adjust to dynamic environments, significantly improving the quality of life for those with hearing difficulties.

Future work

The future of real-time acoustic environment classification holds immense potential to revolutionize the field of hearing aids and assistive auditory systems. To further advance this technology, several key areas warrant exploration. Firstly, expanding the diversity of acoustic environments within datasets, including industrial settings, outdoor spaces, and diverse cultural contexts, will enable models to generalize better to real-world scenarios. Secondly, incorporating advanced feature extraction techniques, such as deep learning-based methods, can capture intricate acoustic patterns and improve classification accuracy. Additionally, exploring ensemble learning, transfer learning, and attention mechanisms can enhance model performance and robustness. Optimizing models for real-time inference on resource-constrained devices, leveraging hardware acceleration, and developing adaptive learning algorithms are crucial for practical implementation. Finally, conducting user-centric evaluations to assess the impact on user experience and personalizing the system to individual needs will ensure its effectiveness and acceptance. By addressing these areas, future research can significantly advance the field, leading to more sophisticated and effective hearing aids and assistive auditory systems that enhance the quality of life for individuals with hearing impairments.

REFERENCES

[1] Student, U. G., et al. "SOUND CLASSIFICATION SYSTEM USING MACHINE LEARNING TECHNIQUES."

[2] Zaheer, Ruba, et al. "A survey on artificial intelligence-based acoustic source identification." IEEE Access 11 (2023): 60078-60108.

[3] Mulimani, Manjunath. Acoustic Scene Classification Using Speech Features. Diss. National Institute of Technology Karnataka, Surathkal, 2020.

[4] LENATTI, M., 2019. Automated detection of hearing loss by machine learning approaches applied to speech-in-noise testing for adult hearing screening.

[5] Gourisaria, Mahendra Kumar, et al. "Comparative analysis of audio classification with MFCC and STFT features using machine learning techniques." Discover Internet of Things 4.1 (2024): 1.

[6] Mohtadifar, Masoud, Michael Cheffena, and Alireza Pourafzal. "Acoustic-and radio-frequency-based human activity recognition." Sensors 22.9 (2022): 3125.

[7] Abdul, Zrar Kh, and Abdulbasit K. Al-Talabani. "Mel frequency cepstral coefficient and its applications: A review." IEEE Access 10 (2022): 122136-122158.

[8] Desai, Nidhi, Kinnal Dhameliya, and Vijayendra Desai. "Feature extraction and classification techniques for speech recognition: A review." International Journal of Emerging Technology and Advanced Engineering 3.12 (2013): 367-371.

[9] Purwins, Hendrik, et al. "Deep learning for audio signal processing." IEEE Journal of Selected Topics in Signal Processing 13.2 (2019): 206-219.

[10] Purwins, Hendrik, et al. "Deep learning for audio signal processing." IEEE Journal of Selected Topics in Signal Processing 13.2 (2019): 206-219.

[11] Doan, Van-Sang, Thien Huynh-The, and Dong-Seong Kim. "Underwater acoustic target classification based on dense convolutional neural network." IEEE Geoscience and Remote Sensing Letters 19 (2020): 1-5.

[12] Abbasi, Ahmed, et al. "A large-scale benchmark dataset for anomaly detection

and rare event classification for audio forensics." IEEE Access 10 (2022): 38885-38894.

[13] Sindhu, Irum, and Mohd Shamrie Sainin. "Automatic Speech and Voice Disorder Detection using Deep Learning-A Systematic Literature Review." IEEE Access (2024).

[14] Sharma, Jivitesh, Ole-Christoffer Granmo, and Morten Goodwin. "Environment sound classification using multiple feature channels and attention based deep convolutional neural network." (2020).

[15] Guzhov, Andrey, et al. "Esresnet: Environmental sound classification based on visual domain models." 2020 25th international conference on pattern recognition (ICPR). IEEE, 2021.