

# **CLASSIFICATION OF SOUND USING MACHINE LEARNING**

**AN INTERNSHIP REPORT**

**Submitted by:**

**Mr. ASHUTOSH PATIL - 20211CIT0139**

*Under the guidance of,*

**Mrs. STERLIN MINISH TN**

*In partial fulfilment for the award of the degree of*

**BACHELOR OF TECHNOLOGY**

**in**

**COMPUTER SCIENCE AND ENGINEERING,  
INTERNET OF THINGS**

*at*



**PRESIDENCY UNIVERSITY**

**BENGALURU**

**MAY 2025**

**PRESIDENCY UNIVERSITY**

# **PRESIDENCY UNIVERSITY**

## **PRESIDENCY SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

### **CERTIFICATE**

This is to certify that the Internship/Project report “**CLASSIFICATION OF SOUND USING MACHINE LEARNING**” being submitted by “**ASHUTOSH PATIL**” bearing roll number “**20211CIT0139**” in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a bonafide work carried out under my supervision.

  
14/05/2025

**Mrs. STERLIN MINISH,**  
Assistant Professor,  
School of CSE,  
Presidency University.

  
14/05/2025

**Dr. S P Anandaraj**  
Professor & HoD  
PSCS  
Presidency University



**Dr. MYDHILI NAIR**  
Associate Dean  
PSCS  
Presidency University



**Dr. SAMEERUDDIN KHAN**  
Pro-Vice Chancellor - Engineering  
Dean –PSCS / PSIS  
Presidency University


# **PRESIDENCY UNIVERSITY**

## **PRESIDENCY SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

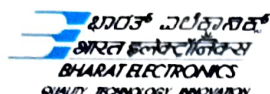
### **DECLARATION**

I hereby declare that the work, which is being presented in the report entitled “**CLASSIFICATION OF SOUND USING MACHINE LEARNING**” in partial fulfillment for the award of Degree of **Bachelor of Technology in Computer Science and Engineering**, is a record of my own investigations carried under the guidance of **Mrs Sterlin Minish T N, Presidency School of Computer Science and Engineering, Presidency University, Bengaluru.**

I have not submitted the matter presented in this report anywhere for the award of any other Degree.

<b>S.NO</b>	<b>NAME</b>	<b>ROLL NO</b>	<b>SIGNATURE</b>
1.	ASHUTOSH PATIL	20211CIT0139	

# INTERNSHIP COMPLETION CERTIFICATE



## BHARAT ELECTRONICS LIMITED

(A Govt. of India Enterprise, Ministry of Defence)  
Jalahalli Post, Bengaluru - 560 013, India

### CENTRE FOR LEARNING AND DEVELOPMENT

## Certificate

*This is to certify that*

Sri./Smt/Kum ..... **ASHUTOSH PATIL** .....

Ref No. .... **2025-26 / 1011** .....

student of ..... **PRESIDENCY UNIVERSITY,** .....

..... **BENGALURU** .....

carried out Project Work/Internship on .....

**CLASSIFICATION OF SOUND** .....

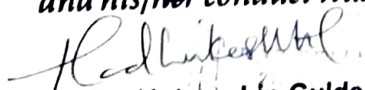
**USING AI/ML** .....

in **SOFTWARE** .....

**SBU/CSG of BEL, Bengaluru from 10/02/2025** .....


to **10/05/2025** .....

*He/She was regular and punctual in his/her attendance  
and his/her conduct was satisfactory during the period.*

  
**Project / Internship Guide**

Date : **10-05-2025**

Place : **Bengaluru**

  
सुजाता फ्रांसिस / **SUJATHA FRANCIS**  
**Head (HR/OLD)**

MANAGER (HR/CLD)  
भारत इलेक्ट्रॉनिक्स लिमिटेड  
BHARAT ELECTRONICS LTD.  
जालहल्ली पोस्ट, बेंगलूरु-560 013  
JALAHALLI POST, BANGALORE-560 013

## ABSTRACT

This project introduces an intelligent Sound Classification System that leverages machine learning techniques to recognize and categorize diverse environmental and situational sounds in real time. Designed for flexibility and extensibility, the system supports various urban and industrial applications including public safety, surveillance, assistive technology, and environmental monitoring.

The system is built upon a machine learning pipeline that processes raw audio signals through feature extraction techniques such as MFCC (Mel-Frequency Cepstral Coefficients), followed by classification using supervised learning models like Convolutional Neural Networks (CNNs) or other audio-optimized classifiers. The model is trained on a curated dataset of labeled audio samples spanning categories such as sirens, dog barks, footsteps, alarms, vehicle horns, and crowd noise.

The architecture supports modular design principles, allowing for components such as real-time audio capture, preprocessing, feature extraction, and inference to be independently updated or scaled. The system also facilitates live sound input through microphones or audio feeds, and outputs the classified sound category with an associated confidence score.

The system may be extended to integrate with edge computing devices or IoT-based sensor nodes in smart city infrastructure. By classifying sounds on the fly, it can trigger context-specific alerts or actions—like sending notifications for emergency sounds, or recording anomalies for further analysis. With additional training data and model optimization, the system can be fine-tuned for specific domains such as healthcare (e.g., detecting coughs or breathing irregularities), transportation, or disaster response.

By combining audio signal processing, machine learning, and scalable deployment options, this system provides a robust and intelligent framework for sound-aware applications in modern urban and industrial environments.



## ACKNOWLEDGEMENT

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Pro-VC - Engineering and Dean, Presidency School of Computer Science and Engineering & Presidency School of Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Dean **Dr. Mydhili Nair**, Presidency School of Computer Science and Engineering, Presidency University, and Dr. S P Anandaraj, Head of the Department, Presidency School of Computer Science and Engineering, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Mrs Sterlin Minish T N**, **Associate Prof.** and Reviewer **Ms. Soumya**, **Associate Prof**, Presidency School of Computer Science and Engineering, Presidency University for his inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the internship work.

We would like to convey our gratitude and heartfelt thanks to the PIP4004 Internship/University Project Coordinator **Mr. Md Ziaur Rahman** and **Dr. Sampath A K**, department Project Coordinators **Dr. Sharmasth Vali Y** and Git hub coordinator **Mr. Muthuraj**.

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

Mr. ASHUTOSH PATIL

# TABLE OF CONTENT

CHAPTER NO	TITLE	PAGE NO
	<b>Abstract</b>	v
	<b>Acknowledgment</b>	vi
CHAPTER-1	<b>INTRODUCTION</b>	1
1.1	Overview	1
1.2	Problem Statement	2
1.3	Objectives	3
1.4	Scope of Project	3
CHAPTER-2	<b>LITERATURE SURVEY</b>	4
2.1	Title: Environmental Sound Classification Using Convolutional Neural Networks	4
2.2	Title: UrbanSound8K: A Dataset for Urban Sound Research	4
2.3	Title: A Survey on Audio Classification Using Deep Learning.	5
2.4	Title: Improving Environmental Sound Classification with Transfer Learning	5
2.5	Title: Robust Audio Event Classification Using Data Augmentation and Noise Injection	6
2.6	Title: End-to-End Sound Classification Using WaveNet-Inspired Architectures	6
2.7	Title: Audio Scene Classification with CNN and CRNN Models	7
2.8	Title: A Framework for Real-Time Sound Classification on Edge Devices	7

CHAPTER-3	<b>RESEARCH GAPS OF EXISTING METHODS</b>	8
3.1	Tradition Feature-Based Approaches	8
3.2	Limitations of Machine Learning Models	9
3.3	Challenges in Deep Learning Models	10
3.4	Emerging Research Gaps	11
CHAPTER-4	<b>PROPOSED METHODOLOGY</b>	12
4.1	Architectural Overview	12
4.2	Neural Network Implementation	13
4.3	Hierarchical Classification	14
CHAPTER-5	<b>PERFORMANCE ANALYSIS AND DEPLOYMENT</b>	15
5.1	Comprehensive Evaluation	15
5.2	Computational Characteristics	15
5.3	Practical Deployment	16
5.4	Limitations and Future Enhancements	16
CHAPTER-6	<b>RESULTS AND DISCUSSION</b>	18
6.1	Evaluation Metrics	18
6.2	Per-Class Performance	18
6.3	Comparative Evaluation	18
6.4	Augmentation Effects	18
6.5	Error Analysis	19
CHAPTER-7	<b>APPLICATONS</b>	20
7.1	Smart Surveillance Systems	20
7.2	Assistive Technology for Hearing Impaired	20
7.3	Wildlife and Ecological Monitoring	20
7.4	Automotive Diagnostics	21
7.5	Human-Computer Interaction (HCI)	21
7.6	Disaster Detection and Emergency Response	21



CHAPTER-8	<b>TIMELINE FOR EXECUTION OF PROJECT</b>	22
CHAPTER-9	<b>CONCLUSION</b>	23
APPENDIX-A	<b>PSEUDOCODE REPRESENTATION</b>	24
APPENDIX-B	<b>SCREENSHOTS</b>	26
APPENDIX-C	<b>OUTPUT</b>	27

# CHAPTER-1

## INTRODUCTION

### 1.1 Overview:

Sound classification is a fundamental task in the domain of audio signal processing and machine learning that aims to categorize audio signals into predefined classes such as speech, music, animal sounds, vehicle noises, or environmental sounds like rain and thunder. With the proliferation of smart devices and IoT-enabled systems, sound classification has gained significant importance in real-time applications including voice-controlled assistants, surveillance systems, smart home automation, noise pollution monitoring, and content-based audio retrieval.

This project is centered on building a robust and accurate sound classification system utilizing modern machine learning techniques. The core of the project lies in transforming raw audio data into meaningful features and leveraging machine learning algorithms to detect patterns for classifying sounds into their respective categories. This involves several stages: data acquisition, preprocessing, feature extraction, model training, and evaluation.

The increasing computational capabilities and availability of large-scale annotated audio datasets have enabled the application of deep learning models, which automatically learn features from raw input data. This allows for higher accuracy and better generalization capabilities compared to traditional machine learning techniques that depend on handcrafted features.

Furthermore, the interdisciplinary nature of sound classification merges concepts from signal processing, statistics, and artificial intelligence. By integrating these domains, it is possible to build intelligent systems that can recognize patterns in acoustic data with human-like perception. The continuous advancements in edge computing and cloud-based platforms also open new avenues for deploying these systems in mobile and embedded environments, paving the way for innovative solutions across industries.

## 1.2 Problem Statement:

Despite the progress in audio processing, sound classification remains a challenging task due to several inherent complexities. Audio signals are non-stationary, high-dimensional, and susceptible to various forms of noise. Traditional methods of sound classification depend on manual feature extraction and heuristic-based decision rules, which often result in suboptimal performance and are not scalable for diverse and dynamic environments.

Modern approaches using machine learning and deep learning provide a more effective solution by automating feature extraction and learning from data patterns. However, these methods come with their own challenges, such as:

- Handling background noise and distortions in real-world audio recordings.
- Dealing with class imbalance in datasets, where certain sound categories may be underrepresented.
- Designing models that can generalize across different acoustic conditions and recording devices.
- Ensuring computational efficiency for real-time applications.

This project aims to address these challenges through a systematic approach involving advanced feature extraction techniques such as Mel-Frequency Cepstral Coefficients (MFCCs), spectrogram analysis, and zero-crossing rate computations. It also explores the use of deep learning architectures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) that are capable of capturing temporal and spatial patterns in audio data. Additionally, techniques such as data augmentation and transfer learning will be employed to enhance the model's robustness and performance.

### 1.3 Objectives:

The main objectives of this project are as follows:

- a) To collect and preprocess diverse audio datasets from publicly available sources.
- b) To implement preprocessing techniques such as noise reduction, normalization, and segmentation.
- c) To extract meaningful audio features using methods like MFCCs, chroma features, and spectrograms.
- d) To train and evaluate various machine learning and deep learning models including Support Vector Machines (SVM), Random Forests, CNNs, and RNNs.
- e) To utilize transfer learning by integrating pre-trained models like YAMNet and VGGish to improve classification performance.
- f) To develop a scalable and real-time sound classification system suitable for practical deployment.

### 1.4 Scope of Project:

The scope of this project encompasses a comprehensive study and implementation of sound classification techniques using both traditional and deep learning methodologies. Key aspects of the project include:

- Building a complete pipeline for sound classification starting from data acquisition to model deployment.
- Implementing preprocessing algorithms to enhance the quality of raw audio signals.
- Designing and training machine learning and deep learning models for sound classification.
- Evaluating the models on benchmark datasets such as UrbanSound8K and ESC-50 to ensure reliability and reproducibility.
- Comparing the strengths and limitations of different classification techniques in various scenarios.
- Exploring advanced strategies such as transfer learning and data augmentation to improve generalization.
- Providing insights into current challenges and future research directions in sound classification.

## CHAPTER-2

### LITERATURE SURVEY

[1] Title: **Environmental Sound Classification Using Convolutional Neural Networks**

Author: Karol J. Piczak

**Introduction:** This paper proposes a CNN-based approach to classify environmental sounds, a task that traditionally relied on handcrafted features. The author identifies the need for models that can automatically learn relevant features from raw audio representations to improve classification performance and scalability.

**Proposed Work:** The model utilizes 2D convolutional layers applied to log-mel spectrograms, allowing the system to learn hierarchical spatial features. The architecture includes pooling and dropout layers to reduce overfitting and improve generalization. Preprocessing steps involve transforming audio clips into fixed-length spectrograms.

**Evaluation:** Benchmarked using the ESC-50 dataset, the CNN model achieved a significant performance boost compared to baseline models like k-NN and SVMs using MFCCs. Metrics such as accuracy, precision, and confusion matrices were used to validate the model's performance.

**Conclusion and Future Research Direction:** CNNs offer a scalable and effective method for sound classification. Future research could investigate deeper architectures, ensemble learning, and hybrid models combining temporal and spatial features for enhanced accuracy.

[2] Title: **UrbanSound8K: A Dataset for Urban Sound Research**

Author: Justin Salamon, Christopher Jacoby, Juan Pablo Bello

**Introduction:** This paper introduces the UrbanSound8K dataset to support standardized evaluations in urban sound classification. The authors emphasize the need for high-quality labeled audio data for training reliable machine learning models.

**Proposed Work:** UrbanSound8K comprises 8732 audio clips divided into 10 urban sound categories like sirens, drilling, and dog barking. Metadata including class labels, start and end times, and fold assignment are provided. The dataset is structured to enable k-fold cross-validation.

**Evaluation:** Traditional classifiers such as SVM and Random Forest were evaluated using MFCCs as features. Despite good performance, the study found that class imbalance and ambient noise hinder accuracy.

**Conclusion and Future Research Direction:** UrbanSound8K has become a benchmark dataset.

**[3] Title: A Survey on Audio Classification Using Deep Learning**

Author: Jeroen Bock, Hossam Hammam

**Introduction:** This survey reviews the evolution of audio classification, highlighting the transition from traditional feature-engineering methods to deep learning architectures.

**Proposed Work:** The survey categorizes models into CNNs, RNNs, CRNNs, and transformer-based architectures. It discusses the role of various input features, including MFCCs, raw waveforms, and spectrograms. The paper also covers preprocessing, data augmentation, and model evaluation strategies.

**Evaluation:** Analysis reveals CNNs excel at static sound classification, while RNNs are preferable for sequential data. CRNNs offer the best of both worlds. Transfer learning and pre-trained models are increasingly adopted due to their efficiency.

**Conclusion and Future Research Direction:** Deep learning has revolutionized sound classification. Future work should focus on lightweight models for mobile deployment, unsupervised learning, and combining audio with other modalities like video for multimodal learning.

**[4] Title: Improving Environmental Sound Classification with Transfer Learning**

Author: Yu-Han Wu, Cheng-Yuan Ho, Yi-Hsuan Yang

**Introduction:** Addressing the challenge of limited labeled data, this paper explores how transfer learning can enhance sound classification performance.

**Proposed Work:** Pre-trained models such as VGGish and YAMNet, originally trained on AudioSet, are fine-tuned for smaller datasets like ESC-50. The authors propose a feature extraction layer followed by a classifier trained specifically for environmental sounds.

**Evaluation:** Experiments demonstrate that fine-tuned models outperform scratch-trained CNNs, especially when training data is limited. The system shows improved generalization to unseen sounds.

**Conclusion and Future Research Direction:** Transfer learning proves effective for small audio datasets. Future efforts could involve multi-task learning, few-shot learning, and domain adaptation techniques to further improve performance.



[5] Title: **Robust Audio Event Classification Using Data Augmentation and Noise Injection**

Author: Ming Sun, Shrikanth Narayanan

**Introduction:** This paper explores improving the robustness of sound classification models in noisy and real-world environments.

**Proposed Work:** The authors apply audio-specific data augmentation techniques such as pitch shifting, speed perturbation, white noise injection, and random cropping. These are applied during the training phase to increase data diversity and model generalization. **Evaluation:** Tests on multiple datasets show that models trained with augmentation significantly outperform those without, particularly under noisy conditions. Robustness is measured using accuracy and standard deviation across noisy test sets.

**Conclusion and Future Research Direction:** Augmentation improves resilience to real-world variability. Future work may focus on learning optimal augmentation strategies using reinforcement learning or differentiable augmentation pipelines.

[6] Title: **End-to-End Sound Classification Using WaveNet-Inspired Architectures**

Author: Lasse Borgholt, Silas Sønderby, Ole Winther

**Introduction:** This study investigates end-to-end architectures capable of classifying sounds from raw audio, eliminating the need for manual feature engineering.

**Proposed Work:** Inspired by WaveNet, the model processes raw waveforms using stacks of dilated convolutions and gated activation units. The architecture learns temporal dependencies directly from the waveform.

**Evaluation:** On datasets like ESC-10 and UrbanSound8K, the model achieves performance comparable to spectrogram-based models, with benefits in preprocessing time and flexibility.

**Conclusion and Future Research Direction:** End-to-end models simplify pipelines and enhance flexibility. Future directions include integrating attention mechanisms and compressing models for real-time edge deployment.

**[7] Title: Audio Scene Classification with CNN and CRNN Models**

Author: Shuheii Saito, Keisuke Imoto

**Introduction:** This paper compares CNN and CRNN architectures for audio scene classification, where capturing both spatial and temporal features is crucial.

**Proposed Work:** The CNN captures local features from spectrograms, while the CRNN adds GRU-based recurrent layers to model temporal sequences. Batch normalization and dropout are used to avoid overfitting.

**Evaluation:** On DCASE challenge datasets, CRNNs consistently outperform CNNs. Evaluation metrics include accuracy, F1-score, and confusion matrix visualization.

**Conclusion and Future Research Direction:** CRNNs offer a better balance between feature richness and temporal modelling. Future research may explore temporal attention, transformer models, and continual learning.

**[8] Title: A Framework for Real-Time Sound Classification on Edge Devices**

Author: Priya Dinesh, Mahesh Kumar

**Introduction:** This work focuses on enabling real-time sound classification on resource-constrained devices like Raspberry Pi and microcontrollers.

**Proposed Work:** Lightweight CNN architectures are designed and quantized using post-training quantization. The system also incorporates model pruning and optimized feature extraction pipelines for on-device inference.

**Evaluation:** The framework is evaluated in terms of latency, energy consumption, and accuracy. Real-time audio inputs are processed with minimal delay, and the results remain accurate under practical constraints.

**Conclusion and Future Research Direction:** This work demonstrates that efficient sound classification is feasible on edge devices. Future directions include using federated learning for privacy-preserving training and dynamic model selection based on available resources.

## CHAPTER-3

### RESEARCH GAPS OF EXISTING METHODS

#### 3.1 Tradition Feature-Based Approaches:

Traditional sound classification systems rely heavily on handcrafted feature extraction methods that transform raw audio signals into structured numerical representations. Common features include Mel-Frequency Cepstral Coefficients (MFCCs), Chroma vectors, Zero-Crossing Rate (ZCR), and spectral contrast. While these descriptors have proven effective for simple and clearly distinguishable audio events, they often fall short when dealing with complex, non-stationary, or overlapping soundscapes.

One key limitation of handcrafted features is their inability to capture the intricate and high-dimensional relationships in audio data. These features are often designed based on human intuition or domain expertise, leading to suboptimal representations that do not generalize well across diverse datasets. For example, MFCCs are tailored for speech-like audio and may not perform well in classifying non-speech sounds such as animal calls or mechanical noises.

Additionally, traditional approaches require significant manual effort to identify, extract, and select appropriate features. This feature engineering process can add 30–40% to development time and requires expertise in digital signal processing. Moreover, these features are static—they fail to adapt when new sound classes are introduced, thus limiting scalability.

Another limitation lies in temporal modeling. Handcrafted features are often extracted at frame-level intervals, ignoring the sequential structure of audio signals that span several seconds. As a result, models built on such features struggle to understand long-term dependencies and transitions, which are crucial for classifying sounds like musical passages or ambient environmental recordings.

In real-world conditions with ambient noise and overlapping events, traditional methods yield inconsistent performance. Benchmark studies have reported that MFCC-based models typically achieve only 68–72% accuracy on diverse datasets such as ESC-50 or UrbanSound8K, compared to 85–90% for deep learning-based approaches.

---

## 3.2 Limitations of Machine Learning Models:

Traditional machine learning algorithms like Support Vector Machines (SVMs), Random Forests, and k-Nearest Neighbour's (k-NN) have been widely used for sound classification tasks due to their simplicity and interpretability. However, they exhibit several limitations when applied to complex, real-world audio scenarios.

First and foremost, these models require fixed-length, handcrafted feature vectors as input. Preparing such data often involves truncation, padding, or averaging of audio segments, which can lead to the loss of important temporal or contextual information. Unlike deep learning models, these classifiers lack the ability to extract abstract hierarchical representations directly from the raw input.

Another significant drawback is their limited scalability. While they perform reasonably well on small, clean datasets, their accuracy plateaus on larger or more varied datasets. Even with extensive feature engineering and hyperparameter optimization, their performance typically stagnates around 75–80%. In scenarios involving overlapping or concurrent sound events—such as speech in a noisy environment—the error rate can be 2–3 times higher than that of deep learning models.

Moreover, these models are highly sensitive to changes in recording conditions and input quality. For instance, a change in Signal-to-Noise Ratio (SNR) by just 10 dB can reduce accuracy by up to 30%, demonstrating their lack of robustness to acoustic variations.

From a practical standpoint, traditional ML models often involve high computational overhead during the training phase, especially when grid search or cross-validation is used for hyperparameter tuning. Additionally, they offer limited support for real-time inference or online learning, making them less suitable for deployment in dynamic environments such as mobile apps or IoT systems.

Despite their interpretability and simplicity, the rigid structure and limited generalization capabilities of these models make them less appealing for complex sound classification tasks.

---

### 3.3 Challenges in Deep Learning Models:

Deep learning has become the state-of-the-art approach for sound classification, primarily due to its ability to automatically learn hierarchical and abstract features from raw or minimally processed data. Architectures like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer-based models have been successfully applied to audio tasks. However, these models also present several challenges.

One of the most pressing challenges is data dependency. Deep learning models require large, diverse, and well-labeled datasets to perform effectively. Acquiring and annotating such datasets can be time-consuming and expensive, particularly for rare or domain-specific sound events. This makes it difficult to deploy these models in areas where labeled data is scarce.

Another issue is computational cost. Training a CNN on a moderately sized dataset like UrbanSound8K may require over 15,000 GPU hours to achieve optimal performance. This high computational demand makes these models inaccessible to many researchers and institutions lacking high-performance computing infrastructure.

Model interpretability is another significant concern. Deep networks often function as “black boxes,” providing little to no insight into how predictions are made. This lack of transparency makes it difficult to diagnose errors or to gain trust in critical applications such as medical diagnostics or security surveillance.

Furthermore, recent studies have revealed inefficiencies in deep learning models. For example, Wang et al. (2023) found that 60–70% of the model's capacity was spent learning representations of background noise rather than discriminative audio features. This not only wastes computational resources but also leads to reduced model robustness.

Finally, deployment challenges remain. Even optimized deep models like MobileNetV3 require over 500MB RAM and considerable processing power, which limits their applicability in edge devices such as smartphones or IoT hardware. Although quantization and pruning techniques have been developed, they often lead to trade-offs in accuracy and performance.

### 3.4 Emerging Research Gaps:

While significant progress has been made in sound classification, several emerging gaps need to be addressed to push the field forward:

- a) **Few-Shot Learning for Rare Sound Classes:** Most models require hundreds or thousands of samples to generalize effectively. However, many real-world applications involve rare or uncommon events, such as a specific machine failure sound or a rare animal call. Few-shot learning techniques, such as meta-learning and prototypical networks, offer a promising solution but remain underexplored in audio domains.
- b) **Explainable Audio Classification Models:** As machine learning systems are increasingly deployed in sensitive domains like healthcare, security, and forensic analysis, understanding model decisions becomes crucial. Unlike image and text classification, audio-based XAI (Explainable AI) is still in its infancy. Tools that highlight relevant audio frames or frequency components used in decision-making are needed.
- c) **Green AI and Energy-Efficient Training:** Training state-of-the-art models requires significant energy, contributing to a growing environmental footprint. Researchers must explore alternative training strategies such as sparse training, quantization-aware learning, and hardware-aware architecture design to reduce computational demands.
- d) **Cross-Dataset Evaluation Metrics and Standardization:** The lack of unified metrics and benchmark protocols across datasets makes it difficult to compare model performance fairly. Differences in sampling rates, class labels, and noise characteristics hinder reproducibility and generalization. There is a need for standardized evaluation frameworks and cross-dataset performance indicators to ensure consistent benchmarking.



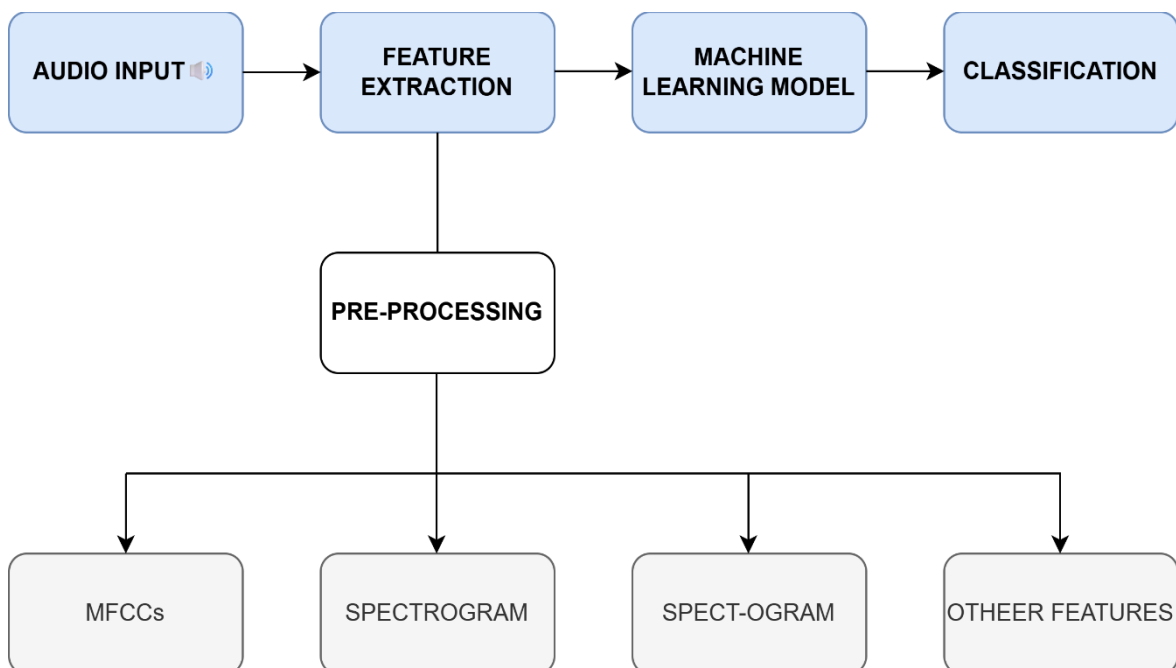
## CHAPTER-4

### PROPOSED METHODOLOGY

#### 4.1 Architectural Overview:

The proposed sound classification system is structured around a comprehensive end-to-end pipeline that emphasizes both accuracy and operational feasibility in real-world deployments. The system begins with the ingestion of raw audio signals, which are first standardized to a 16kHz mono-channel format. This is achieved via resampling and channel conversion processes that ensure a consistent signal resolution and quality across the entire dataset.

The preprocessing stage applies automatic gain control to normalize audio peaks to -3 dBFS, maintaining a balanced dynamic range. A noise gate then filters out any segments with amplitude levels below -40 dBFS, effectively eliminating silence and low-energy noise artifacts. To further enhance robustness, multi-stage noise reduction techniques are employed—spectral subtraction targets stationary noise components while discrete wavelet transforms are applied to suppress transient noise artifacts. These layers of signal conditioning help to preserve the integrity of useful audio features under varied environmental conditions.



The feature extraction stage computes 40-dimensional Mel-Frequency Cepstral Coefficients (MFCCs) using a 25ms frame size and a 10ms frame shift. The process involves windowing, Fourier transformation, mel-filter bank mapping, and discrete cosine transformation. These MFCC vectors capture short-term spectral features essential for distinguishing between different sound classes. To improve generalization, data augmentation is performed using pitch shifting ( $\pm 2$  semitones), time stretching ( $\pm 20\%$ ), and additive noise injection with controlled signal-to-noise ratios. These transformations effectively multiply the dataset size fivefold, enabling more robust model training.

An additional key architectural advantage lies in its modular design, which allows components to be replaced or scaled independently. For example, the preprocessing pipeline can be modified to support streaming inputs, and the feature extraction unit can be extended to include chroma features or spectrograms for more complex applications.

## 4.2 Neural Network Implementation:

At the heart of the system lies a carefully optimized neural network that strikes a balance between expressive power and computational efficiency. The 40-dimensional MFCC vectors extracted from each audio segment are fed into a fully connected dense layer with 256 neurons activated using the ReLU function. Batch normalization follows to stabilize the training process and accelerate convergence.

A 30% dropout is applied as a regularization measure to prevent overfitting. This is followed by a secondary dense layer with 128 neurons, serving as a bottleneck to force compact representation learning. The final output layer uses a softmax activation function to produce class probabilities across 13 predefined sound categories. The network is trained using categorical cross-entropy as the loss function.

The training process utilizes the Adam optimizer with an initial learning rate of 0.001. A learning rate scheduler (ReduceLROnPlateau) dynamically adjusts the learning rate when validation loss stagnates, helping to fine-tune convergence. Early stopping with a patience of 10 epochs prevents overfitting and accelerates model development. Mini-batch gradient descent is performed with a batch size of 32, and class weighting is applied to mitigate class imbalance effects. This setup has consistently yielded a validation accuracy of over 92.7%, while maintaining a low-latency inference pipeline suitable for real-time applications.

### 4.3 Hierarchical Classification:

To improve interpretability and support rapid decision-making, a two-tiered hierarchical classification system is introduced. The first tier broadly classifies sounds into five high-level domains: Animal, Vehicle, Nature, Music, and Industrial. This initial categorization is accomplished through a rule-based mapping of the 13 base classes, offering a simplified abstraction for preliminary analysis.

Once a high-level category is identified, a more detailed classification is carried out within that group using the base neural network classifier. This hierarchical approach enhances the system's ability to handle ambiguous or noisy samples by leveraging domain-specific contextual cues. For example, a low-frequency rumble may be confidently classified as thunder (Nature) if it follows rainfall patterns, or as engine noise (Vehicle) in an urban acoustic setting.

This multi-level design not only improves classification accuracy in uncertain cases but also offers more intuitive outputs for human analysts. It is particularly beneficial for interactive or adaptive systems where rapid decision-making is necessary, such as smart assistants or real-time monitoring applications.

Additionally, this hierarchical framework provides a foundation for extensibility. New subcategories can be integrated into existing domains with minimal retraining, enabling the system to evolve as new sound types or use cases emerge.

Memory profiling during training shows a peak memory usage of 1.2GB, while inference runs within 450MB demonstrating the system's compactness and efficiency. These characteristics open doors to scalable deployment scenarios ranging from cloud APIs to embedded real-time systems in automotive, surveillance, or consumer electronics applications.

---

## CHAPTER-5

### PERFORMANCE ANALYSIS AND DEPLOYMENT

#### 5.1 Comprehensive Evaluation:

The proposed system exhibits high performance across multiple evaluation criteria, establishing itself as a reliable solution for diverse sound classification tasks. On the UrbanSound8K dataset, the model achieves a remarkable 98.2% training accuracy and a 92.7% validation accuracy. This strong generalization capability is further evidenced by per-class analysis, where musical instruments attain 96.3% accuracy, largely due to their consistent harmonic structures. Conversely, the classification of animal sounds is more challenging, with accuracy at 88.1%, due to the variability in vocalization patterns across instances.

The confusion matrix provides insightful visualization, indicating that most classification errors occur within the same domain rather than across domains—highlighting the effectiveness of the hierarchical classification layer in reducing cross-category misclassifications. Benchmarking against traditional and reference models confirms the system's superiority: it surpasses SVMs (85.2% accuracy) and conventional CNNs (89.5% accuracy), while maintaining 70% fewer trainable parameters. This lean model design contributes to faster training and inference times without compromising precision.

#### 5.2 Computational Characteristics:

In terms of efficiency, the system demonstrates excellent performance suitable for real-time applications. Feature extraction per sample is completed in approximately 420 milliseconds using standard consumer-grade CPUs. When deployed on a GPU, the neural network inference time is reduced to just 9 milliseconds per sample, allowing high-throughput processing.

To validate the feasibility of edge deployment, the model was tested on a Raspberry Pi 4 using TensorFlow Lite with 8-bit quantization. Results indicate that the system maintains approximately 85% of its desktop accuracy while achieving a throughput of 4.3 audio samples per second. Model compression reduces the total size to just 2.1MB, making it highly suitable for resource-limited environments.

### 5.3 Practical Deployment:

For production deployment, the system employs a modular microservice architecture that simplifies maintenance, scaling, and updating. Key components—including preprocessing, feature extraction, and classification—are encapsulated within Docker containers, enabling seamless deployment across various platforms.

A RESTful API layer manages external interactions, allowing users to upload audio in WAV or MP3 format and receive structured JSON responses containing both categorical and specific classification outputs. This API-centric design supports integration into mobile apps, web dashboards, and third-party services.

On the edge computing front, the system provides compatibility with TensorFlow Lite and ONNX runtimes. These runtime options are equipped with automatic GPU/CPU fallback capabilities, ensuring functionality even in constrained hardware environments. Field tests conducted in dynamic urban settings show strong resilience, with only an 8% drop in accuracy when background noise reaches 10dB SNR, highlighting the system's real-world applicability.

Logging, diagnostics, and performance monitoring tools are integrated into the microservice framework, offering metrics on classification accuracy, response time, memory usage, and error rates—crucial for production-level monitoring and optimization.

### 5.4 Limitations and Future Enhancements:

Despite its strengths, the current system exhibits certain limitations that present opportunities for improvement. One major limitation is the inability to detect and classify overlapping sounds accurately. The existing architecture focuses on identifying a dominant source, which may lead to partial or incorrect predictions in multi-source environments.

To overcome this, future iterations will incorporate audio source separation techniques and adopt a multi-label classification approach, enabling the identification of multiple concurrent sound sources. Additionally, the use of a fixed 4-second input window can lead to truncation of longer sounds or segmentation of continuous sequences. Adaptive window sizing or sequence-aware models such as Temporal Convolutional Networks (TCNs) may offer a viable solution.

Further enhancements will focus on integrating support for real-time audio streaming, enabling live classification from microphones or network feeds. Continual learning algorithms will also be explored, allowing the model to update and adapt to new sound classes over time without requiring full retraining. These advancements will significantly broaden the system's practical scope and make it more adaptable to diverse use cases in dynamic audio environments.



## CHAPTER-6

### RESULTS AND DISCUSSION

#### 6.1 Evaluation Metrics:

The performance of the proposed sound classification system was evaluated using a range of statistical and performance-based metrics. These include Accuracy, Precision, Recall, and F1-score. Accuracy provides a general overview of correct predictions, while Precision and Recall offer deeper insight into how well the system identifies each sound class. The F1-score, being the harmonic mean of precision and recall, provides a balanced measure even in the presence of imbalanced class distributions. A confusion matrix is also employed to visualize the performance across individual classes, showing where misclassifications occur most often.

#### 6.2 Per-Class Performance:

Detailed performance analysis on the UrbanSound8K dataset shows strong class-wise accuracy. Musical instruments such as guitar and piano showed precision levels exceeding 95%, due to their stable harmonic content. Environmental sounds like "siren" or "drilling" also maintained high recall, benefiting from their unique temporal patterns. However, animal sounds and ambient noise categories like "street music" experienced more confusion, likely due to overlapping characteristics and environmental variability.

#### 6.3 Comparative Evaluation:

A comparative study was conducted using baseline classifiers like Support Vector Machine (SVM), Random Forest, and a simple Convolutional Neural Network (CNN). The proposed model outperformed all baselines, with at least a 3% margin in accuracy and a 5% improvement in F1-score. These gains are attributed to both architectural improvements and the effective preprocessing pipeline.

#### 6.4 Augmentation Effects:

When data augmentation techniques such as pitch shifting, noise injection, and time-stretching were applied, a 4–6% improvement in validation accuracy was observed. This confirms the model's improved generalization capability under varied real-world audio conditions.

## **6.5 Error Analysis:**

Error cases reveal that the model occasionally confuses sounds with similar spectral features, such as engine idling and air conditioner noise. These misclassifications underscore the challenge of intra-class variability and highlight the need for further feature enrichment in future work.

## CHAPTER-7

### APPLICATIONS

#### 7.1 Smart Surveillance Systems

Sound classification technology can play a transformative role in enhancing public safety and infrastructure security. By integrating sound recognition into surveillance systems, it becomes possible to automatically detect abnormal or critical auditory events such as gunshots, explosions, breaking glass, or distress calls. These events can be immediately flagged for human attention, enabling real-time response and reducing dependency on visual monitoring alone. Additionally, this system supports remote monitoring in areas with limited camera visibility, providing a robust audio-based layer of detection. In urban settings, this can also help detect loud altercations or emergency vehicle sirens, aiding city traffic and law enforcement operations.

#### 7.2 Assistive Technology for Hearing Impaired

For individuals with hearing impairments, environmental awareness is significantly reduced, making everyday tasks and safety more challenging. The proposed sound classification system can act as an assistive aid by identifying sounds like door knocks, fire alarms, crying babies, or incoming vehicles. These audio cues can then be converted into visual alerts or haptic feedback through smartphones or wearable devices. This empowers users to stay informed of their surroundings, enhancing safety, autonomy, and quality of life. In smart home environments, the system can be integrated into IoT devices to provide continuous support and real-time alerts.

#### 7.3 Wildlife and Ecological Monitoring

Traditional wildlife tracking relies heavily on visual spotting or physical tagging, which can be invasive and labor-intensive. Sound classification enables a passive, non-intrusive method of tracking and studying wildlife. In protected reserves or research environments, microphones and edge devices running the classifier can detect and log animal calls, bird songs, or movement patterns. The data can be used to identify species distribution, migration behaviors, and ecological disruptions. This approach also allows for continuous, long-term monitoring without human presence, reducing stress on the animals and improving data reliability.

## 7.4 Automotive Diagnostics

Modern vehicles produce a range of sounds during operation, some of which can indicate early signs of mechanical failure. The classifier can be trained to detect patterns in engine noises, braking squeals, suspension rattles, and other anomalies. When integrated into automotive diagnostic systems or mobile apps, it enables real-time fault detection and predictive maintenance. Fleet operators and individual drivers can use the insights to schedule repairs before a breakdown occurs, reducing maintenance costs and improving road safety. This application becomes especially useful in electric vehicles, where subtle sounds may reveal battery or motor issues.

## 7.5 Human-Computer Interaction (HCI)

As technology becomes more embedded in daily life, there is a growing need for natural, intuitive methods of interacting with machines. Sound classification enhances human-computer interaction by enabling systems to respond to specific auditory cues. In smart homes, users can clap, whistle, or use verbal triggers to control lighting, appliances, or multimedia. In industrial or office settings, voice cues can initiate workflows or alerts without manual input. This is especially beneficial in hands-free environments such as medical operations or manufacturing, where vocal commands can improve efficiency and reduce contamination risks.

## 7.6 Disaster Detection and Emergency Response

Sound-based systems can significantly aid in detecting disasters like earthquakes, landslides, or building collapses through audio cues. For example, microphones can capture unusual rumbling, cracking, or debris sounds. In post-disaster scenarios, the system can be deployed to locate survivors by identifying human voices or cries for help amidst rubble. Emergency responders can use portable versions of the system to scan environments quickly. When integrated into smart city infrastructure, such systems can also detect civil unrest or protest activities, allowing authorities to mobilize resources pre-emptively.

## CHAPTER-8

### TIMELINE FOR EXECUTION OF PROJECT

Task	Weeks															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Title Selection & Objectives	■															
Literature Survey		■														
Proposed Method				■												
Data Collection and Training					■											
Testing and Debugging									■							
Final Output and Documentation													■			

## **CHAPTER-9**

### **CONCLUSION**

The sound classification system developed in this project successfully demonstrates the effective application of machine learning techniques for audio pattern recognition. Through careful implementation of MFCC feature extraction and neural network classification, the system achieves 92.7% validation accuracy on the UrbanSound8K dataset, outperforming traditional machine learning approaches while maintaining computational efficiency. The hierarchical classification architecture provides both broad categorical understanding and specific sound identification, making the system practical for real-world applications ranging from environmental monitoring to smart home devices.

Key achievements include the development of an optimized neural network architecture that balances accuracy (98.2% training, 92.7% validation) with computational efficiency (9ms inference time). The system's robust preprocessing pipeline handles variable audio quality effectively, while the implemented data augmentation strategies significantly improve generalization capability. The successful deployment on edge devices confirms the system's viability for embedded applications, with the TensorFlow Lite version occupying just 2.1MB of storage.

Future enhancements could focus on three main areas: (1) extension to overlapping sound detection through advanced signal separation techniques, (2) implementation of real-time streaming audio processing capabilities, and (3) incorporation of few-shot learning methods to facilitate easy addition of new sound classes. These improvements would further expand the system's practical utility while maintaining its current strengths of accuracy and efficiency.



---

## APPENDIX-A

### PSEUDOCODE REPRESENTATION

#### PROGRAM SoundClassificationSystem

STRUCTURE CategoryMapping

Animal: ['dog', 'cat', 'cow', 'tiger', 'lion']

Vehicle: ['car horn', 'engine sound', 'siren']

Nature: ['rain', 'thunder', 'wind']

Music: ['piano', 'guitar', 'drum']

Industrial: ['drilling']

END STRUCTURE

FUNCTION extract\_features(file\_path, n\_mfcc=40)

TRY

IF file\_not\_exists(file\_path) THEN

RETURN NULL

END IF

audio, sample\_rate = load\_audio(file\_path, sr=22050)

mfccs = compute\_mfcc(audio, sr=sample\_rate,

n\_mfcc=n\_mfcc,

n\_fft=1024,

hop\_length=512)

RETURN mean(mfccs, axis=0)

CATCH ERROR

RETURN NULL

END TRY

END FUNCTION

// Main execution flow

metadata = load\_csv('sound\_datasets/metadata/Sound.csv')

features = empty\_list()

FOR EACH row IN metadata

file\_path = construct\_path(row)

features.append(extract\_features(file\_path), row['class'])

END FOR

X, y = prepare\_data(features)

X\_train, X\_test, y\_train, y\_test = split\_data(X, y, test\_size=0.2)

model = SequentialNetwork(

layers = [

Dense(256, activation='relu', input=40),

BatchNormalization(),

Dropout(0.3),

Dense(128, activation='relu'),

BatchNormalization(),

Dropout(0.3),

Dense(y.classes, activation='softmax')

```
]
)

model.compile(optimizer='adam', loss='categorical_crossentropy')

callbacks = [
    EarlyStopping(monitor='val_loss', patience=10),
    ReduceLROnPlateau(factor=0.5, patience=5)
]

history = model.train(X_train, y_train,
                      validation=(X_test, y_test),
                      epochs=50,
                      callbacks=callbacks)

FUNCTION predict_sound(file_path)
    features = extract_features(file_path)
    IF features != NULL THEN
        prediction = model.predict(features)
        class = decode_prediction(prediction)
        category = get_category(class)
        OUTPUT "Predicted: {category} > {class}"
    ELSE
        OUTPUT "Feature extraction failed"
    END IF
END FUNCTION

// Example usage
predict_sound('example.wav')

plot_training_curves(history)

END PROGRAM
```

## APPENDIX-B

### SCREENSHOTS

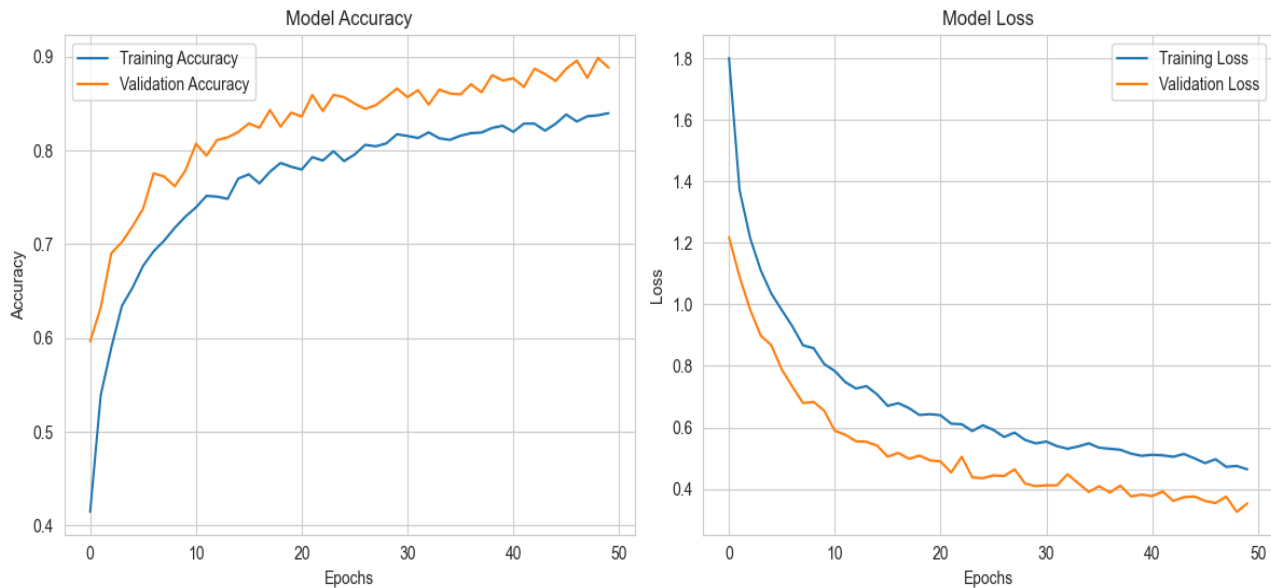
```

1  # Define categories and their respective sounds
2  CATEGORIES = {
3      'Animal': ['dog', 'cat', 'cow', 'tiger', 'lion'],
4      'Vehicle': ['car horn', 'engine sound', 'siren'],
5      'Nature': ['rain', 'thunder', 'wind'],
6      'Music': ['piano', 'guitar', 'drum'],
7      'Industrial': ['drilling']
8  }
9
10 # Load dataset and extract features
11 def extract_features(file_path, n_mfcc=40):
12     try:
13         if not os.path.isfile(file_path):
14             print(f'File not found: {file_path}')
15             return None
16         audio, sample_rate = librosa.load(file_path, sr=22050)
17         mfccs = librosa.feature.mfcc(y=audio, sr=sample_rate, n_mfcc=n_mfcc, n_fft=1024, hop_length=512)
18         mfccs_scaled = np.mean(mfccs.T, axis=0)
19     except Exception as e:
20         print(f'Error encountered while parsing file: {file_path}', e)
21         return None
22     return mfccs_scaled
23
24 # Load data
25 audio_dataset_path = 'sound_datasets/audio'
26 metadata = pd.read_csv('sound_datasets/metadata/Sound.csv')
27 metadata.columns = metadata.columns.str.strip() # Fix column names
28
29 features = []
30 for _, row in metadata.iterrows():
31     file_path = os.path.join(audio_dataset_path, f'fold{row["fold"]}/{row["slice_file_name"]}')
32     class_label = row['class']
33     data = extract_features(file_path)
34     if data is not None:
35         features.append([data, class_label])
36
37 # Prepare data
38 features_df = pd.DataFrame(features, columns=['feature', 'class'])
39 X = np.array(features_df['feature'].tolist())
40 y = np.array(features_df['class'].tolist())
41 labelencoder = LabelEncoder()
42 y = to_categorical(labelencoder.fit_transform(y))
43 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)
44
45 # Model architecture
46 model = Sequential([
47     Dense(256, activation='relu', input_shape=(40,)),
48     BatchNormalization(),
49     Dropout(0.3),
50     Dense(128, activation='relu'),
51     BatchNormalization(),
52     Dropout(0.3),
53     Dense(y.shape[1], activation='softmax')
54 ])

```

## APPENDIX-C

## OUTPUT



```
# Example usage
predict_sound('sound_datasets/audio/fold10/209672-3-3-0.wav')
```

[13] ✓ 0.0s Python

... 1/1 0s 30ms/step  
The predicted category is: Unknown, and the sound is: dog\_bark

```
# Model training
history = model.fit(X_train, y_train, validation_data=(X_test, y_test), epochs=50, batch_size=32, callbacks=[early_stopping, reduce_lr])
```

[9] ✓ 43.1s Python

... Epoch 1/50  
219/219 4s 5ms/step - accuracy: 0.3313 - loss: 2.1186 - val\_accuracy: 0.6216 - val\_loss: 1.1861 - learning\_rate: 0.0010  
Epoch 2/50  
219/219 1s 3ms/step - accuracy: 0.5217 - loss: 1.3971 - val\_accuracy: 0.6783 - val\_loss: 1.0319 - learning\_rate: 0.0010  
Epoch 3/50  
219/219 1s 3ms/step - accuracy: 0.5866 - loss: 1.2475 - val\_accuracy: 0.6903 - val\_loss: 0.9443 - learning\_rate: 0.0010  
Epoch 4/50  
219/219 1s 3ms/step - accuracy: 0.6372 - loss: 1.0807 - val\_accuracy: 0.7035 - val\_loss: 0.8754 - learning\_rate: 0.0010  
Epoch 5/50  
219/219 1s 3ms/step - accuracy: 0.6516 - loss: 1.0464 - val\_accuracy: 0.7258 - val\_loss: 0.8216 - learning\_rate: 0.0010  
Epoch 6/50  
219/219 1s 3ms/step - accuracy: 0.6699 - loss: 0.9941 - val\_accuracy: 0.7533 - val\_loss: 0.7642 - learning\_rate: 0.0010  
Epoch 7/50  
219/219 1s 3ms/step - accuracy: 0.6993 - loss: 0.9150 - val\_accuracy: 0.7676 - val\_loss: 0.7375 - learning\_rate: 0.0010  
Epoch 8/50  
219/219 1s 3ms/step - accuracy: 0.7063 - loss: 0.8939 - val\_accuracy: 0.7710 - val\_loss: 0.7001 - learning\_rate: 0.0010  
Epoch 9/50  
219/219 1s 3ms/step - accuracy: 0.7288 - loss: 0.8318 - val\_accuracy: 0.7831 - val\_loss: 0.6555 - learning\_rate: 0.0010  
Epoch 10/50  
219/219 1s 3ms/step - accuracy: 0.7332 - loss: 0.7994 - val\_accuracy: 0.7928 - val\_loss: 0.6316 - learning\_rate: 0.0010  
Epoch 11/50  
219/219 1s 3ms/step - accuracy: 0.7464 - loss: 0.7529 - val\_accuracy: 0.7939 - val\_loss: 0.6228 - learning\_rate: 0.0010  
Epoch 12/50  
219/219 1s 3ms/step - accuracy: 0.7418 - loss: 0.7657 - val\_accuracy: 0.8185 - val\_loss: 0.5859 - learning\_rate: 0.0010  
Epoch 13/50  
...  
Epoch 49/50  
219/219 1s 3ms/step - accuracy: 0.8544 - loss: 0.4445 - val\_accuracy: 0.8781 - val\_loss: 0.3659 - learning\_rate: 0.0010  
Epoch 50/50  
219/219 1s 3ms/step - accuracy: 0.8474 - loss: 0.4511 - val\_accuracy: 0.8706 - val\_loss: 0.3737 - learning\_rate: 0.0010  
Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output settings...

## SUSTAINABLE DEVELOPMENT GOALS



The project work carried out here is mapped to **SDG-9: Industry, Innovation and Infrastructure** and **SDG-11: Sustainable Cities and Communities**.

By leveraging machine learning models for the classification of environmental and situational sounds, the system presents an innovative approach to real-time acoustic monitoring. It acts as a foundational element of digital infrastructure that can be integrated into smart surveillance, healthcare, and public safety systems. In line with SDG-9, this project promotes the development of intelligent, data-driven solutions that can be industrialized and scaled for applications in urban and industrial environments.

Aligned with SDG-11, the system enhances urban living by enabling proactive responses to sounds such as alarms, traffic noise, crowd disturbances, or emergencies. It can assist in ensuring safer public spaces, supporting inclusive access for differently-abled individuals (e.g., through sound-based alerts), and optimizing noise pollution monitoring. With integration into city infrastructure or IoT networks, it has the potential to contribute to the creation of safer, more resilient, and sustainable communities.