ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

SONICSENTINEL: AI-POWERED GUNFIRE EVENT RECOGNITION FROM AMBIENT SOUNDSCAPE

BOYINA GOPI RAJU[1], GAVINI MEKALA NAGA MANIKANTA[2]

[1] HOD &Assoiate Professor, Department of CSE, Chirala Engineering College, Chirala, Andhra Pradesh, India-523157.

[2] MCA Student, Department of CSE, Chirala Engineering College, Chirala, Andhra Pradesh, India-523157.

ABSTRACT

Gunshot sound recognition has emerged as an important research area due to the increasing need for intelligent surveillance, public safety monitoring, and automated threat detection systems. Traditional gunshot detection methods often rely on manual monitoring or basic acoustic sensing techniques, which are limited in their ability to accurately distinguish firearm sounds from other environmental noises. Variations in recording conditions, background interference, and similarities between different weapon sounds further increase the complexity of reliable firearm identification. These challenges necessitate the development of an automated and intelligent system capable of accurately classifying gunshot sounds in real-world environments. To address this problem, the proposed system introduces a Machine Learning (ML)-based gunshot sound classification framework that analyzes firearm audio recordings and categorizes them into predefined weapon classes. The system performs audio preprocessing and extracts discriminative acoustic features such as Mel-Frequency Cepstral Coefficients (MFCC), chroma features, spectral contrast, and zero-crossing rate using the Librosa library. The extracted features are then utilized for classification through multiple ML algorithms, including Gaussian Naïve Bayes Classifier (GNBC), Logistic Regression (LR), Linear Discriminant Analysis Classifier (LDAC), and Extra Trees Classifier (ETC). Performance evaluation is conducted using accuracy, precision, recall, F1-score, classification reports, and confusion matrices to identify the most effective classification approach. A user-friendly graphical interface is developed using the Tkinter framework to facilitate dataset management, feature extraction, model training, evaluation, and firearm prediction. The system supports role-based authentication for administrators and users, enabling secure access to various functionalities.

KEYWORDS: Gunshot Sound Recognition, Machine Learning, Audio Classification, MFCC, Firearm Detection, Acoustic Feature Extraction, Intelligent Surveillance, Threat Detection.

1. INTRODUCTION

1.1 Overview

Global firearm-related violence has escalated in recent years, amplifying the demand for rapid, AI-based gunshot detection systems [1]. These technologies analyse acoustic signatures frequency, amplitude, and temporal waveforms to differentiate various firearm discharges in real time, Fig. 1.1 as shown. AI-enabled systems offer law enforcement a non-invasive, automated alternative to traditional forensic ballistics, delivering location-based alerts for immediate response. In 2022, the United States experienced over 650 mass shootings defined as incidents involving four or more victims, excluding the shooter with total firearm-related deaths surpassing 43,000. This alarming trend has driven urban centers to adopt advanced acoustic surveillance. ShotSpotter, deployed in more than 100 U.S. cities (including Chicago, New York, and San Francisco), utilizes sensor networks to triangulate gunfire and dispatch alerts in under a minute. Independent studies report up to 97% system accuracy in controlled conditions, though real-world performance varies. Estimated annual deployment and maintenance costs can exceed 65,000 per square mile [2,3].

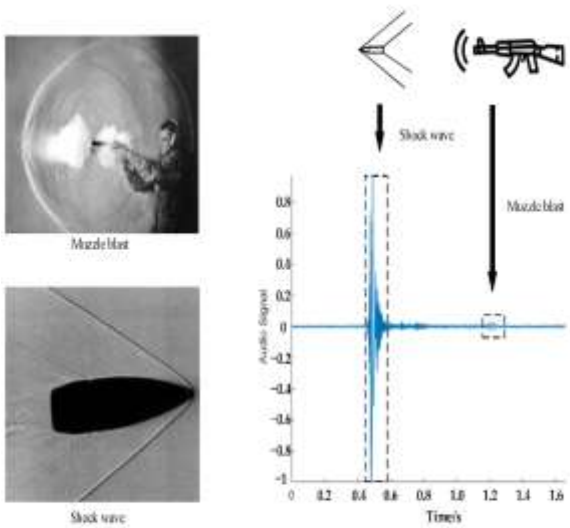

Fig. 1.1: Gunshot Event Recognition

Research on ShotSpotter reveals mixed outcomes. Some evaluations note faster police response and improved evidence collection including assistance in incidents not reported via 911 but criminological analyses suggest no statistically significant reduction in violent crime or firearm homicides. Critics also raise concerns over false positives, racial profiling, and high infrastructure costs, leading some cities to terminate contracts amid concerns over their efficacy [4,5]. Internationally, regions in Latin America face severe gun violence. Brazil, for example, recorded nearly 38,800 homicides in 2024 an intentional homicide rate of 17.9 per 100,000 with firearms implicated in the majority of cases [6]. Despite being the country's lowest rate in over a decade, gun violence remains widespread, driven by organized crime, firearm accessibility, and social inequality [8,9,10].

1.2 Problem Definition

Many existing monitoring environments depend on basic detection systems or human supervision to identify important sound events, which limits accuracy and responsiveness. Traditional approaches often rely on threshold-based sound detection or simple alarms that cannot differentiate between harmless environmental noise and potentially dangerous acoustic signals. This leads to frequent false alarms or missed detections, reducing the reliability of monitoring systems. Environmental audio is inherently complex because it contains overlapping signals such as traffic, conversations, machinery, and weather noise, making meaningful interpretation challenging. Conventional systems lack the capability to analyse these layered signals effectively or extract meaningful acoustic characteristics from raw sound data. As a result, organizations face difficulty in obtaining precise information from audio streams, which affects their ability to maintain safe and secure environments. The problem therefore lies in the absence of efficient analytical mechanisms capable of understanding and classifying real-world sound patterns accurately under diverse conditions.

1.3 Research Motivation

In modern urban environments, ensuring public safety has become increasingly complex due to population growth, rising security concerns, and the expansion of smart infrastructure. Real-time companies operating in domains such as surveillance technology, smart city analytics, transportation security, and industrial monitoring generate enormous volumes of sensor data every second. These organizations rely heavily on data analysis to interpret environmental signals and detect unusual activities that could indicate potential threats or emergencies. Without analytical processing, raw audio or sensor data remains unstructured and unusable for decision-making. Companies need intelligent systems that can automatically process incoming data streams, identify patterns, and highlight critical events requiring attention. Manual monitoring is not only inefficient but also prone to fatigue and human error, which can result in missed alerts. Therefore, data-driven analytical solutions are essential for transforming continuous environmental inputs into actionable insights that enhance operational awareness and response readiness.

1.4 Significance

The significance of this research lies in its contribution to the advancement of intelligent monitoring technologies that can strengthen safety infrastructure across various sectors. Environmental audio analysis plays an important role in improving situational awareness by enabling systems to interpret sound signals that may indicate unusual or risky conditions. Such analytical capabilities support organizations in making informed decisions quickly and effectively, especially in environments where rapid response is critical. This research also contributes to the broader field of data-driven surveillance, which focuses on converting raw sensory inputs into structured and meaningful information. By highlighting the importance of automated interpretation of environmental data, the study emphasizes how modern analytical techniques can enhance reliability, efficiency, and accuracy in monitoring systems. Furthermore, it demonstrates the value of integrating computational intelligence into real-world safety applications, ultimately supporting the development of smarter, more responsive environments that can adapt to dynamic conditions.

2. LITERATURE SURVEY

1. **Nijhawan et al. (2022)** proposed a transformer-based deep learning model for firearm identification using gunshot audio recordings. The study utilized attention mechanisms to analyze acoustic signatures of different weapons and achieved high classification accuracy, demonstrating the potential of artificial intelligence in enhancing public safety and surveillance systems.
2. **Li et al. (2022)** developed a fast gunshot type identification framework based on knowledge distillation techniques. Their approach reduced computational complexity while maintaining reliable recognition performance, making it suitable for real-time gunshot detection applications in resource-constrained environments.
3. **Kabealo et al. (2023)** introduced a comprehensive multi-firearm and multi-orientation gunshot audio dataset containing recordings from various weapon categories under different recording conditions. The dataset provides a valuable benchmark for training and evaluating machine learning models for firearm sound recognition and localization tasks.
4. **Irungu et al. (2023)** investigated the use of transfer learning for gunshot detection in urban acoustic environments. By leveraging pre-trained audio classification models, the study successfully differentiated gunshots from environmental noises and demonstrated improved detection performance in complex real-world soundscapes.
5. **Sigmund (2024)** proposed an effective gunshot detection method based on the short-term entropy of signal energy. The technique was specifically designed to identify weak firearm sounds in noisy environments and showed enhanced robustness in challenging surveillance scenarios where conventional detection methods often fail.

3. PROPOSED METHODOLOGY

3.1 Overview

The proposed methodology defines a structured and analytical framework for developing an intelligent environmental audio recognition system capable of identifying critical acoustic events under real-world conditions. The system follows a systematic pipeline beginning with audio data acquisition, preprocessing, and feature extraction, followed by multi-model training, evaluation, and prediction. Multiple classical ML algorithms are integrated to ensure robust classification performance and comparative analysis. A lightweight storage mechanism is used to store extracted features and trained models, enabling efficient reuse and faster execution. Visualization modules provide waveform analysis and performance graphs to support interpretability and monitoring, in fig. 4.1. The framework also supports reusable model loading, reducing computation time and enabling rapid deployment. This structured architecture ensures reliability, scalability, and adaptability in handling continuous environmental audio streams.

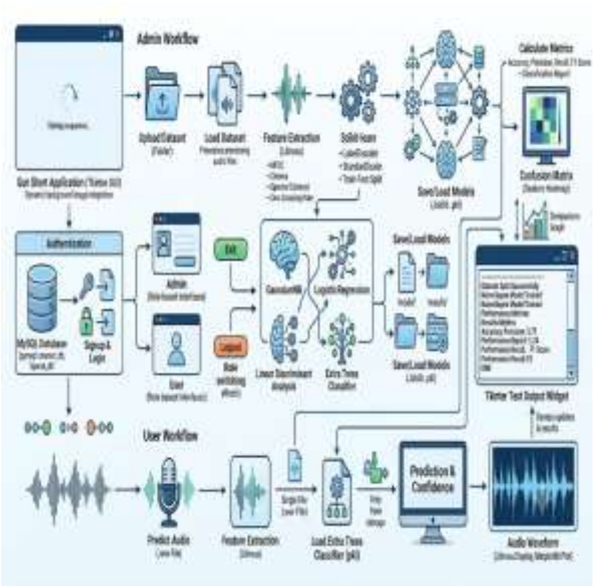

Fig. 3.1: Proposed System Architecture.

3.2 Preprocessing

The preprocessing stage acts as the foundation of the system because it transforms raw environmental audio signals into structured numerical data suitable for ML models. Since real-world audio contains noise, variations in amplitude, silence segments, and recording inconsistencies, direct use of raw signals would lead to inaccurate predictions. Therefore, this stage systematically reads each audio file, standardizes its format, extracts meaningful acoustic characteristics, and converts them into consistent feature vectors. The objective of preprocessing is to remove irrelevant variations while preserving distinctive sound patterns that help classifiers differentiate between classes. This process ensures that the learning algorithms receive clean, uniform, and informative inputs for training and prediction. Without this stage, model performance would be unstable and unreliable.

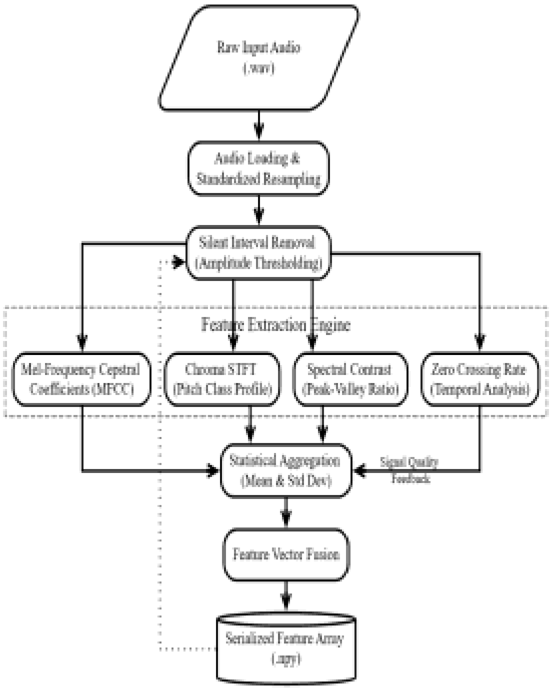

Fig. 3.2: Internal workflow of Preprocessing.

Audio File Loading: Fig. 3.2 presents, The system begins by reading each sound file from the dataset directory using an audio processing library. The waveform signal and sampling rate are extracted while maintaining original sound quality. Loading raw waveforms allows for the analysis of frequency, amplitude, and temporal variations, ensuring all samples are in a standardized digital format.

Signal Normalization and Noise Handling: After loading, the audio signal is prepared to ensure consistency across different recordings. This stage manages variations in length, amplitude, and recording quality to minimize the impact of recording conditions on classification results. This creates uniformity across samples and ensures a fair comparison between sound instances.

Randomization and Reproducibility Control: The dataset is randomly shuffled to eliminate ordering bias and ensure representative class distribution across both subsets. A fixed random seed value is applied during this process to ensure reproducibility, allowing for consistent data division across multiple experimental runs for accurate performance comparison.

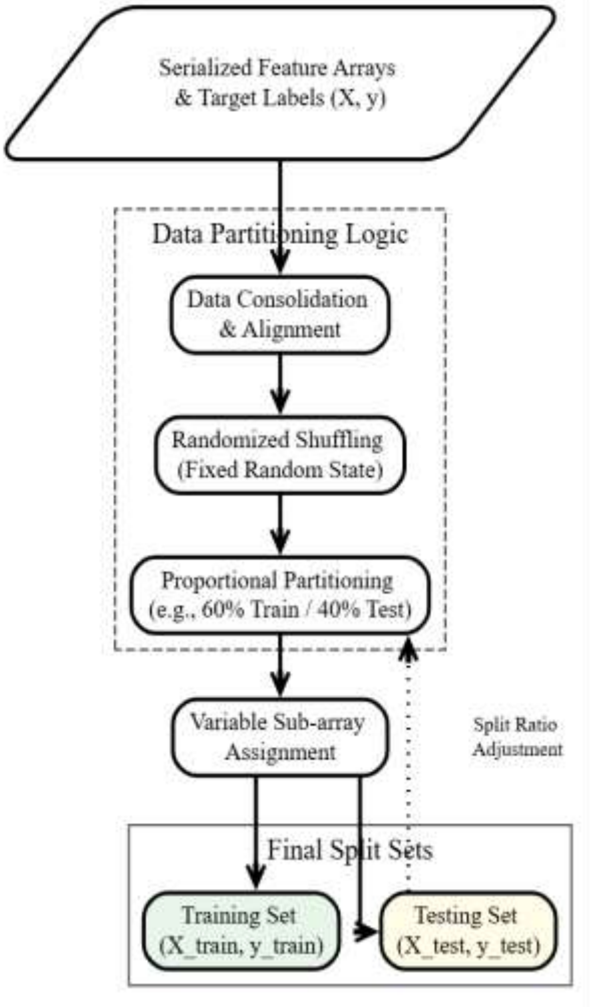

Fig. 3.3: Internal workflow of train test split.

Applying the Splitting Function: A dedicated splitting function is used to automatically divide the dataset while maintaining the correspondence between features and labels. This process generates four distinct outputs: training features, testing features, training labels, and testing labels, which form the core foundation for model development.

Training and Testing Data Allocation: The training subset is passed to ML algorithms to adjust internal parameters and build decision boundaries. Conversely, the testing subset is kept strictly separate, serving as unseen input to simulate real-world scenarios and provide an unbiased estimate of the system's practical deployment accuracy.

Validation and Training Readiness: Following the split, the system verifies data integrity and ensures class distributions remain balanced. This validation step prevents runtime errors and confirms that the dataset structure is consistent. Once validated, the separated sets are forwarded to the modelling pipeline for algorithm fitting and evaluation.

3.4 Model Building &Training

The model building and training phase forms the core of the proposed gunshot sound classification system, where extracted and standardized audio feature vectors are utilized to learn discriminative patterns associated with different firearm types. After completing preprocessing and train–test splitting, multiple ML models are constructed to analyse the high-dimensional acoustic features. Existing baseline classifiers, namely GNBC, LR, and LDAC, are trained to establish reference performance levels and to understand the behaviour of linear and probabilistic learning approaches on gunshot audio data. In parallel, a more robust ensemble-based ETC is implemented as the proposed model to effectively handle feature complexity, randomness, and non-linear decision boundaries. Each model is trained using the training dataset, with learned parameters optimized to minimize classification error. Trained models are persistently stored using Joblib, allowing reuse without repeated training. This structured model building and training strategy ensures fair comparison among classifiers, improves generalization, and lays a strong foundation for accurate firearm identification from unseen gunshot audio samples.

3.4.1 Gaussian Naïve Bayes Classifier (GNBC)

The GNBC stage focuses on learning statistical patterns from the training dataset and using probability theory to classify environmental audio signals into their respective categories. This algorithm assumes that the extracted features follow a Gaussian distribution and that each feature contributes independently to the prediction. During training, it estimates statistical parameters such as mean and variance for every feature within each class. During testing, it applies Bayes' theorem to compute the probability that a given audio sample belongs to each class and selects the class with the highest probability. This method is computationally efficient and works well for high-dimensional feature vectors like audio descriptors. The operation combines statistical modelling with probabilistic decision-making to perform classification reliably. It serves as one of the baseline models for evaluating system performance.

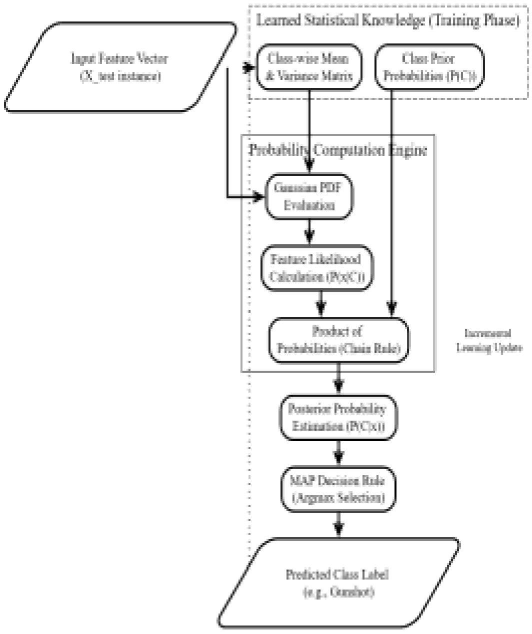

Fig. 3.4: Internal workflow of GNBC.

Likelihood and Posterior Probability Calculation: When a test sample is received, the model evaluates its feature vector against the learned Gaussian distributions. It calculates the likelihood the probability that the sample's features match a specific class and combines this with the prior probability using Bayes' Theorem. This results in a posterior probability for each class given the observed features.

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$

Classification Output and Prediction: The classifier selects the category with the highest posterior probability and outputs the predicted class label. This label represents the acoustic category whose statistical pattern most closely matches the test sample's features. These predictions are then compared against actual labels to generate accuracy and performance metrics.

3.4.2 Logistic Regression (LR)

The LR operates as a supervised learning algorithm that models the relationship between extracted audio features and their corresponding class labels using a probabilistic linear decision boundary. Instead of directly predicting classes, it estimates the probability that a given input belongs to each category and then selects the class with the highest probability. During training, the algorithm learns optimal weights for each feature so that the predicted probabilities closely match actual labels. It uses an optimization process to minimize prediction error and improve classification accuracy. Because it is based on a linear model, it performs efficiently even with high-dimensional feature vectors. The process combines mathematical optimization with probability estimation to produce reliable classification decisions. This classifier serves as an important baseline model for performance comparison.

Probability Transformation (Sigmoid Function): The linear output is passed through a sigmoid function, mapping the numerical value into a probability range between 0 and 1. This transformation allows predictions to be interpreted as likelihoods, where values closer to 1 indicate higher confidence in a class. This step defines the model as a probabilistic classifier.

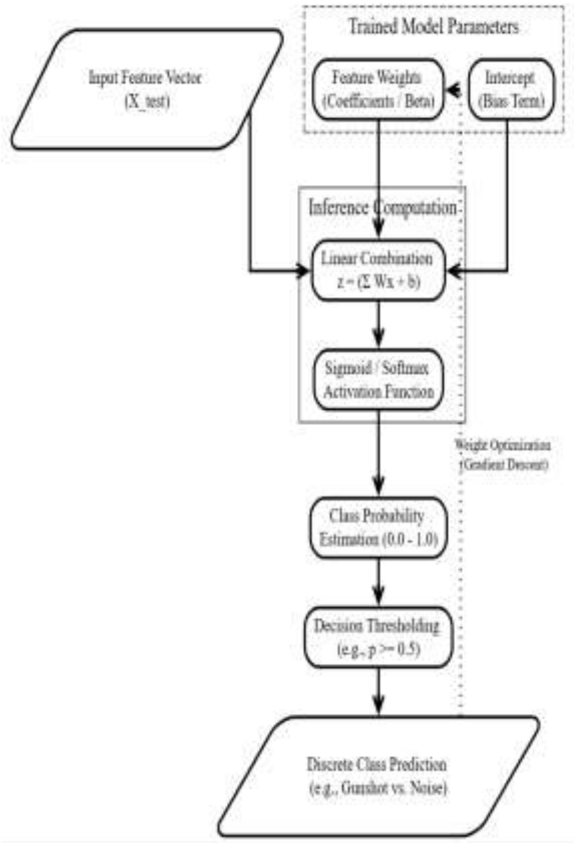

Fig. 3.5: Internal workflow of LR.

Error Measurement and Optimization: The predicted probability is compared with the actual class label using a loss function to quantify the prediction error. To minimize this loss, the algorithm iteratively adjusts feature weights based on their contribution to the error. This optimization process continues across multiple cycles, gradually refining the model's accuracy.

3.4.3 Linear Discriminant Analysis classifier (LDAC)

The LDAC works by finding a mathematical transformation that separates different classes as clearly as possible based on their feature values. Instead of directly predicting classes from raw features, it first researches the data into a new space where class differences become more distinguishable. It assumes that data from each class follows a normal distribution and that all classes share a similar covariance structure. During training, it calculates statistical properties of each class and determines optimal boundaries that maximize separation between them. During testing, it maps new samples into this transformed space and assigns them to the most probable class. This approach combines dimensionality reduction with classification, making it both efficient and effective for structured datasets.

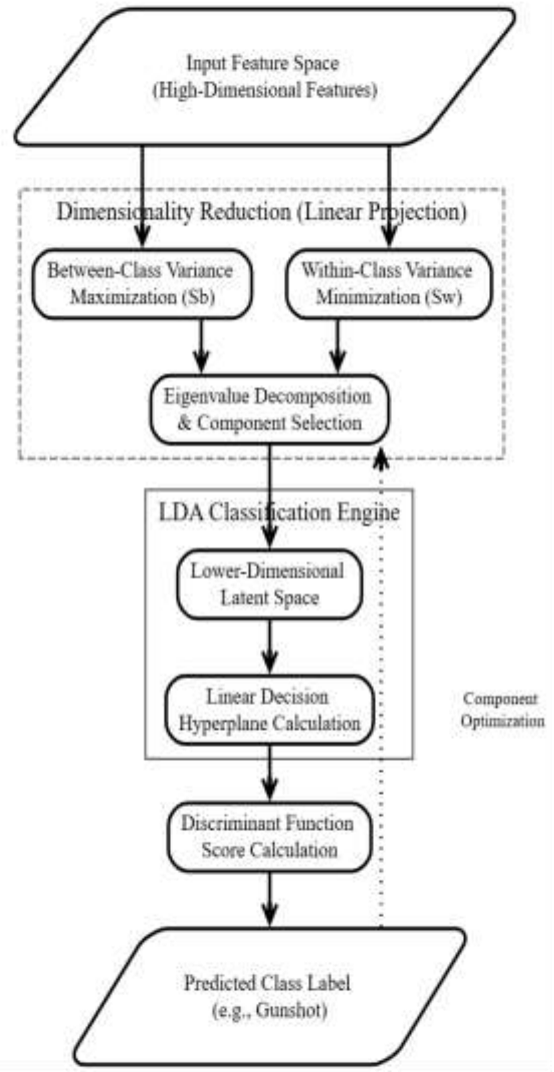

Fig. 3.6: Internal workflow of LDAC.

Optimal Projection and Feature Transformation: Using the scatter values, the model determines a projection direction that maximizes class separation by solving an optimization problem. The goal is to find a transformation that spreads class centers apart while keeping samples of the same class close together. The original feature vectors are then projected onto these new discriminant axes, producing a transformed dataset with pronounced class separation and reduced dimensional complexity.

Final Classification Output: The predicted label is returned as the final output, representing the category that best matches the sample's feature characteristics. This prediction serves as the basis for performance evaluation or real-time detection, completing the classification pipeline.

3.3.4 Extra Trees Classifier (ETC)

The ETC operates as an ensemble learning algorithm that builds multiple decision trees and combines their predictions to produce a final classification result. Instead of relying on a single decision structure, it generates many randomized trees that collectively analyse feature patterns and vote for the most likely class. This randomness helps reduce overfitting and improves generalization performance. During training, the model constructs numerous trees using randomly selected features and random split thresholds, which increases diversity among trees. During testing, each tree independently predicts a class label, and the final output is determined through majority voting.

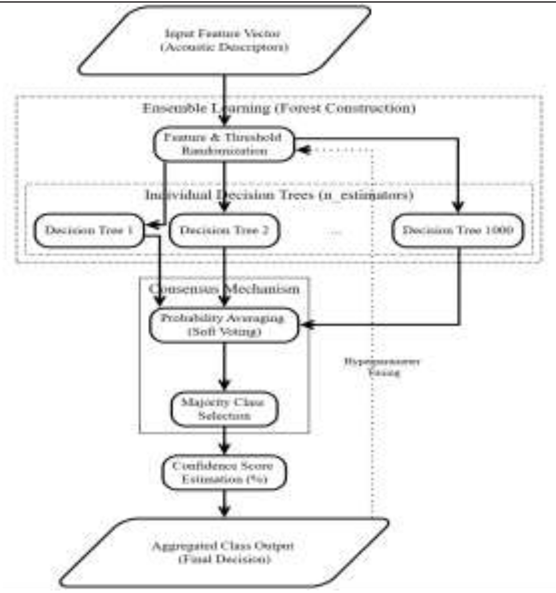

Fig. 3.7: Internal workflow of ETC.

Training Data and Ensemble Initialization: Fig. 3.7, The classifier first accepts the training feature matrix and label vector produced after preprocessing and dataset splitting. Before training begins, the algorithm defines configuration parameters such as the number of trees, maximum depth, and random state. These settings control the forest's complexity and computational cost, preparing the model structure for the learning phase.

Aggregation and Final Output Generation: All individual tree predictions are combined using a majority voting mechanism, where each tree casts one vote for a predicted class. The class receiving the most votes is selected as the final output. This aggregation ensures a robust and stable classification, as the consensus decision of the ensemble is less prone to errors than any single-tree method.

Advantages of ETC

- Provides high classification accuracy because it combines predictions from many decision trees instead of relying on a single model, which improves overall reliability.
- Reduces overfitting effectively due to random feature selection and random split thresholds, allowing the model to generalize well to unseen data.

4. RESULTS AND DISCUSSION

4.1 Results description

The results section presents the performance evaluation of the implemented audio classification system using multiple ML algorithms. Each model is assessed based on standard evaluation metrics such as accuracy, precision, recall, and F1-score to measure its effectiveness in recognizing different sound categories. The analysis highlights how well the system distinguishes between classes under real-world audio variations. Comparative results are used to identify the most reliable classifier for environmental sound recognition. Visualization tools such as graphs and confusion matrices further support interpretation of model behaviour and prediction patterns.

Fig. 4.1: Data Uploading and Features extraction.

Fig. 4.1 illustrates the initial stage of the gunshot sound classification system, where the audio dataset is uploaded through the Tkinter-based graphical user interface. After selecting the dataset folder, the system automatically reads the audio files and extracts relevant acoustic features such as MFCCs, chroma features, spectral contrast, and zero-crossing rate using the Librosa library. The extracted features are converted into numerical feature vectors suitable for machine learning models. The GUI displays the processing status, dataset information, and feature extraction results, enabling users to monitor the preprocessing stage efficiently.

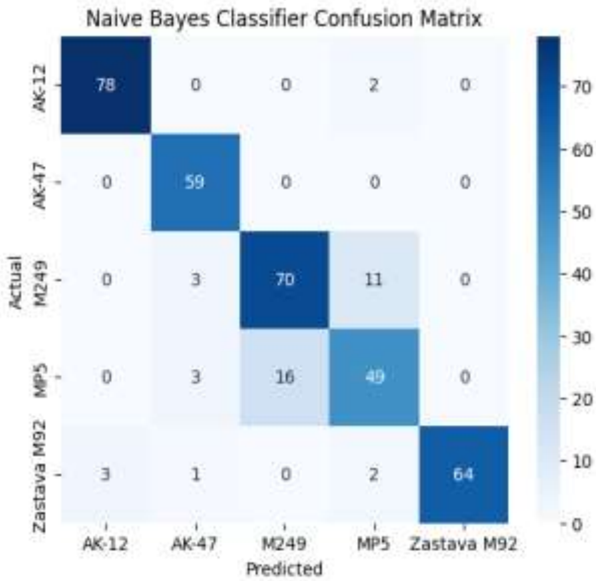

Fig. 4.2: Confusion matrix obtained for GNBC.

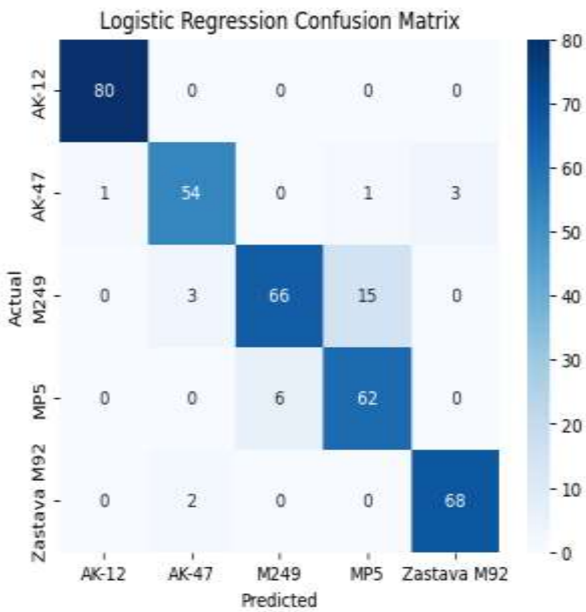

Fig. 4.3: Confusion matrix obtained for LR.

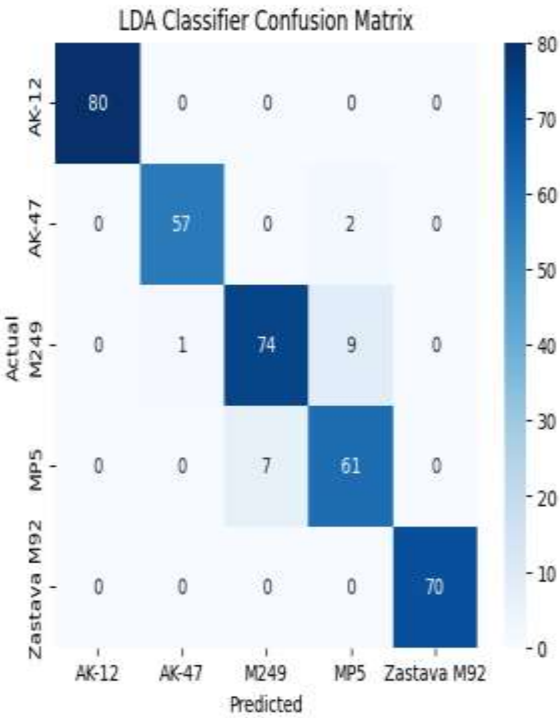

Fig. 4.4: Confusion matrix obtained for LDAC.

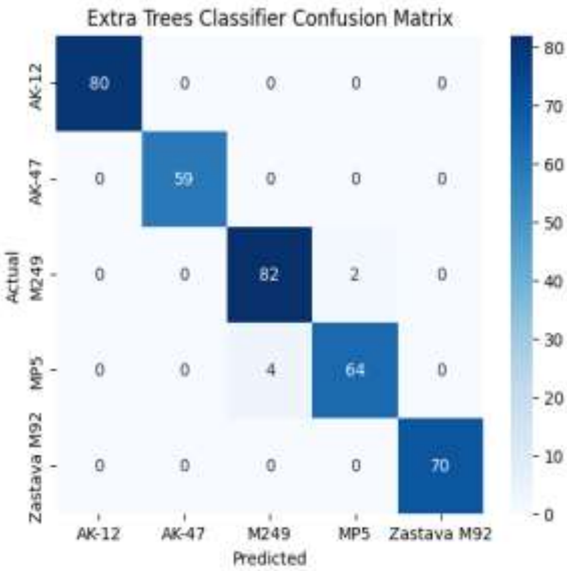

Fig. 4.5: Confusion matrix obtained for ETC.

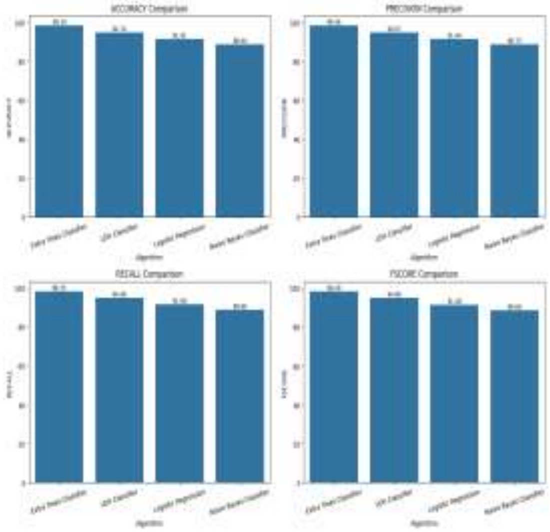

Fig. 4.6: Overall models comparison.

5. CONCLUSION

This paper successfully developed an automated gunshot sound classification system using audio signal processing and machine learning techniques. Acoustic features such as MFCCs, chroma features, spectral contrast, and zero-crossing rate were extracted from gunshot audio recordings to effectively represent the characteristics of different firearm sounds. Multiple classification algorithms, including Gaussian Naïve Bayes Classifier (GNBC), Logistic Regression (LR), and Linear Discriminant Analysis Classifier (LDAC), were implemented and evaluated to establish baseline performance. To improve classification accuracy and robustness, an Extra Trees Classifier (ETC) was proposed and employed as the final prediction model. Experimental results demonstrated that the ETC achieved superior performance compared to the other classifiers, providing reliable identification of firearm categories from audio recordings. A user-friendly Tkinter-based GUI was also developed to facilitate dataset management, model training, evaluation, and real-time prediction of unseen gunshot audio samples. The integration of machine learning techniques with an interactive GUI makes the system practical and accessible for real-world applications. Overall, the proposed solution provides an efficient, scalable, and accurate framework for automatic gunshot sound classification, with potential applications in public safety, surveillance, law enforcement, and security monitoring systems.

REFERENCES

[1] The Global Burden of Disease 2016 Injury Collaborators. Global Mortality from Firearms, 1990–2016. JAMA 2018, 320, 792–814.

[2] Werbick, M.; Bari, I.; Paichadze, N.; Hyder, A.A. Firearm Violence: A Neglected 'Global Health' Issue. Glob. Health 2021, 17, 120.

[3] Raza, A.; Rustam, F.; Mallampati, B.; Gali, P.; Ashraf, I. Preventing Crimes Through Gunshots Recognition Using Novel Feature Engineering and Meta-Learning Approach. IEEE Access 2023, 11, 103115–103131.

[4] Zadey, S.; Branas, C.C.; Morrison, C.N. Gun Violence: A Global Problem in Need of Local Solutions. Lancet 2024, 403, 2783–2784.

[5] Priya, M.; Shendre, S.; Pati, P. Enhanced Gunshot Sound Detection Using AlexNet and XGBoost from Fourier Spectrograms. In Proceedings of the 4th International Conference on Intelligent Technologies (CONIT), Bangalore, India, 21–23 June 2024; pp. 1–7.

[6] Valliappan, N.; Pande, S.; Vinta, S.; Reddy, S. Enhancing Gun Detection with Transfer Learning and YAMNet Audio Classification. IEEE Access 2024, 12, 58940–58949.

[7] Chen, Z.; Zhu, P.; Qiu, W.; Guo, J.; Li, Y. Deep Learning in Automatic Detection of Dysphonia: Comparing Acoustic Features and Developing a Generalizable Framework. Int. J. Lang. Commun. Disord. 2022, 58, 279–294.

[8] Tibrewal, N.; Leeuwis, N.; Alimardani, M. Classification of Motor Imagery EEG Using Deep Learning Increases Performance in

Inefficient BCI Users. PLoS ONE 2022, 17, e0268880.

[9] Chen, Z.; Zheng, H.; Wu, L.; Huang, J.; Yang, Y. Deep-Transfer-Learning-Based Intelligent Gunshot Detection and Firearm Recognition Using Tri-Axial Acceleration. IEEE Internet Things J. 2025, 12, 5891–5900.

[10] Zhang, Y.; Hu, X.; Ma, J.; Wang, X.; Luo, H.; Wu, Z.; Zhang, S.; Shi, D.; Yu, Y.; Qiu, X.; et al. Clinical Applicable AI System Based on Deep Learning Algorithm for Differentiation of Pulmonary Infectious Disease. Front. Med. 2021, 8, 753055.