International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

GUNSHOT SOUND IDENTIFICATION USING CONVOLUTIONAL NEURAL NETWORKS AND SPECTROGRAM FEATURES

***Shakira banu L*[1], *Deepak rajalingam B*[2] , *Bharathi dhasan B*[3]**

Assistant Professor[1] , UG Student[2,3,4]

Department of Artificial Intelligence And Data Science, Dhanalakshmi Srinivasan Engineering College Perambalur, Tamil Nadu, India

saravanan.s@gmail. com[1], akshayasanthosh2802@gmail.com[2],malavikaanandan1@gmail.com[3],limajohnson949@gmail.com[4]

ABSTRACT:

In recent years, the rise in gun-related violence and mass shootings has created an urgent demand for intelligent surveillance systems capable of identifying and responding to gunfire incidents in real time. This project proposes a machine learning-based solution to detect and classify gunshot sounds using audio analysis and deep learning techniques. The objective is to accurately identify whether a sound is a gunshot or not, and in some cases, further classify the type of firearm used, thereby enabling timely alerts and enhancing public safety.The system begins by collecting a diverse set of audio samples, including both gunshot and non-gunshot sounds, primarily sourced from online platforms like YouTube. These audio files are converted into .wav format and segmented into uniform 2-second clips to standardize the dataset. In the preprocessing stage, noise reduction and normalization techniques are applied to ensure clean and consistent audio input. To capture meaningful features from these audio samples, the Mel-Frequency Cepstral Coefficients (MFCCs) are extracted. MFCCs are widely recognized for their ability to represent the timbral and spectral characteristics of sound, making them highly effective for distinguishing between different audio classes.Once the features are extracted, they are fed into a Convolutional Neural Network (CNN), which is designed to learn spatial hierarchies in the data. The CNN model is trained on a labeled dataset and optimized using metrics like accuracy, precision, and recall. After successful training, the model is capable of classifying incoming audio as either gunshot or non-gunshot with high reliability. The model's performance is rigorously evaluated using confusion matrices and graphical tools to ensure its practical effectiveness in real-world scenarios.To make the system accessible to end-users, a web-based interface is developed using HTML, CSS, Bootstrap, and Flask. This interface allows users to upload audio files and instantly receive classification results. The backend is powered by a Python server that integrates the trained CNN model, enabling smooth and responsive interaction between the user and the AI system.Overall, this project presents a robust and scalable solution for real-time gunshot detection using deep learning and audio processing techniques. It not only demonstrates the potential of combining MFCC features with CNNs for sound classification but also offers a practical application that can aid in crime prevention, emergency response, and public security systems.

Keywords: Gunshot Detection, Convolutional Neural Network (CNN), MFCC, Audio Classification, Intelligent Surveillance, Deep Learning.

INTRODUCTION:

Gun violence is a growing concern across the globe, demanding the development of intelligent surveillance technologies that can detect and respond to gunfire events swiftly. Traditional gunshot detection systems rely on acoustic sensors and rule-based or classical machine learning approaches that analyze sound properties like frequency and intensity. These systems, however, often struggle in real-world environments with high ambient noise and variable acoustic conditions.Recent developments in artificial intelligence, especially deep learning, have introduced more powerful tools for audio classification. Convolutional Neural Networks (CNNs), known for their success in image processing, can also be applied to audio by converting sound signals into spectrograms or other visual feature representations like Mel-Frequency Cepstral Coefficients (MFCCs). MFCCs capture key characteristics of audio signals that are relevant to human auditory perception, making them well-suited for gunshot detection.This study proposes a novel system that combines CNNs and MFCCs to classify gunshot audio from other sounds with high accuracy. The model is trained on a diverse dataset collected from open sources and is capable of identifying different types of firearms based on their unique sound signatures. The system is designed to be scalable, cost-effective, and ready for real-time deployment in surveillance and law enforcement settings.

EXISTING SYSTEM:

Traditional gunshot detection systems operate using arrays of microphones placed in strategic urban locations. These microphones capture ambient audio, which is analyzed using rule-based logic or machine learning algorithms. Detection decisions are typically based on acoustic features like sound amplitude, frequency bandwidth, duration, and peak intensity.

Additionally, machine learning classifiers such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Forests (RF) have been used to improve classification accuracy. However, these models depend on handcrafted feature sets, which limits their ability to generalize across different acoustic environments or gun types.These systems often require high-fidelity sensors and centralized processing infrastructure, increasing deployment cost. Moreover, their performance is hindered in noisy settings or during overlapping sound events. As a result, the reliability and scalability of existing systems remain a significant concern, especially for widespread deployment in smart cities or mobile surveillance units.

DRAWBACKS:

The limitations of conventional systems pose serious challenges to effective gunshot detection. These drawbacks include:

- **Manual Feature Engineering**: Feature extraction in classical models is a manual process, requiring domain expertise and limiting adaptability to new types of gunfire or noise patterns.
- **Low Detection Accuracy**: The inability to capture complex sound characteristics leads to misclassifications, especially in acoustically dense environments.
- **High False Positives**: Background noises such as fireworks, thunder, or construction sounds often trigger false alarms, reducing system credibility.
- **Environmental Sensitivity**: Systems degrade in performance with changes in weather, crowd density, or recording distance.
- **High Cost of Infrastructure**: Implementation of traditional acoustic sensor networks demands specialized hardware and significant maintenance overhead.
- **Scalability Constraints**: These systems are difficult to extend to large areas or mobile platforms without significant additional investment.

These shortcomings necessitate a more intelligent, robust, and flexible approach—one that the proposed system aims to address.

PROPOSED SYSTEM :

To overcome the identified challenges, this research proposes a deep learning-based gunshot classification system that leverages the power of CNNs and MFCCs. The key components and methodology are as follows:

- **Data Acquisition**: Gunshot audio clips are collected from open-access datasets such as YouTube and Kaggle. The dataset includes different gun types (e.g., AK-47, M16, MP5), cleaned and segmented into 2-second clips.
- **Feature Extraction using MFCC**: The raw audio files are processed into MFCCs, which provide a compact and information-rich representation of the sound. These features capture both frequency and time-domain properties that are crucial for identifying gunshot signatures.
- **CNN Architecture**: A custom CNN model is designed to take MFCC inputs. It consists of multiple convolutional layers for feature extraction, pooling layers for dimensionality reduction, and dense layers for classification.
- **Regularization Techniques**: Batch Normalization and Dropout layers are used to improve generalization and avoid overfitting.
- **Multi-class Classification**: The final layer uses Softmax activation to classify input sounds into multiple gun categories, enabling identification of the weapon type.
- **Real-Time Processing**: The trained model can classify live or uploaded audio in real time and trigger appropriate alerts or messages.

This system automates feature learning, improves detection robustness, and offers scalability without relying on expensive infrastructure.

ADVANTAGES :

The proposed system offers several significant benefits over traditional methods:

- **High Accuracy**: Deep learning enables superior recognition of complex audio patterns, resulting in better detection performance.
- **Real-Time Capability**: Efficient processing and inference make it suitable for live deployment in surveillance systems.
- **Low False Positive Rate**: Enhanced feature extraction reduces the likelihood of confusing gunshots with similar sounds.
- **Weapon Type Classification**: The system can identify specific gun types, providing valuable insights during investigations.
- **Cost-Effective**: No need for specialized hardware or complex sensor networks—standard audio recordings are sufficient.
- **Scalable Architecture**: Easily extendable to different environments, whether fixed, mobile, or cloud-based platforms.

SYSTEM ARCHITECTURE:

The architecture of the system is divided into several key modules:

1. ***Audio Acquisition***
    - Gunshot sounds are gathered from online datasets and cleaned using audio editing tools like WavePad.
    - Audio clips are converted to WAV format, segmented into uniform lengths, and labeled by gun type.
2. ***Preprocessing and MFCC Extraction***
    - Each audio clip is processed to extract MFCC features, which represent the energy distribution across different

frequency bands.

3. ***CNN Model Training***
   - The MFCCs are passed to a CNN, which learns discriminative patterns for classification.
   - The architecture includes convolutional, pooling, and fully connected layers, along with Dropout for regularization.
4. ***Classification and Output***
   - The trained model performs multi-class classification to identify the type of gunfire.
   - If a gunshot is detected, the system raises an alert, logs the data, and can optionally send a notification.

This modular architecture ensures flexibility and adaptability for different deployment scenarios.

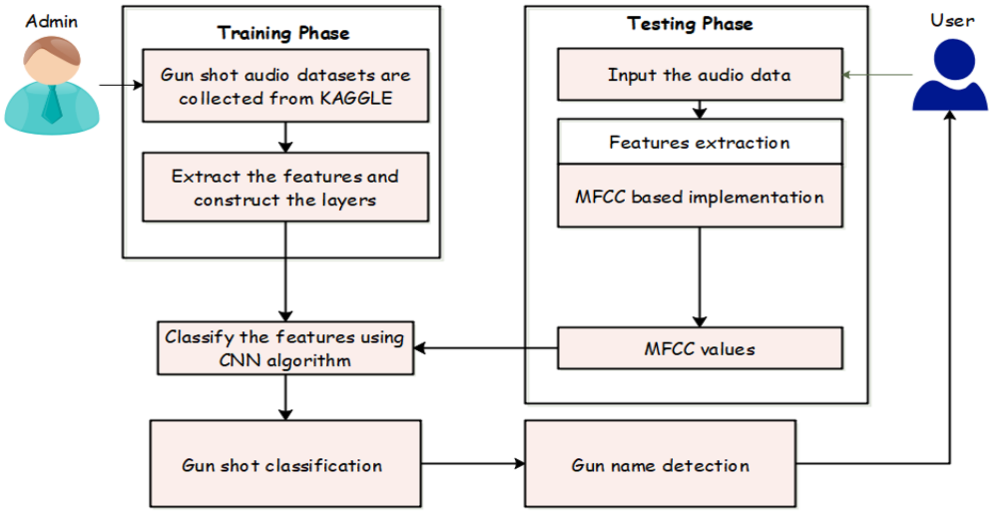

Fig 1. System Architecture

LIST OF MODULES :

1. Audio Acquisition Module
2. Feature Extraction using MFCC Module
3. CNN Model Training Module
4. Gunshot Classification Module
5. Performance Evaluation Module
6. Web Interface Module

MODULE DESCRIPTION :

1. AUDIO ACQUISITION MODULE

The **Audio Acquisition Module** forms the foundation of the system by collecting gunshot and non-gunshot audio clips from online sources such as YouTube. These clips are downloaded, converted into `.wav` format using tools like FFmpeg, and segmented into consistent 2-second audio samples. The segmented clips are then organized into labeled directories based on their sound type, such as different firearm categories or ambient noises. This structured and labeled dataset is critical for training a reliable machine learning model.

2. FEATURE EXTRACTION USING MFCC MODULE

The **Feature Extraction using MFCC Module** transforms the raw audio signals into numerical representations that can be processed by the CNN. It specifically uses Mel-Frequency Cepstral Coefficients (MFCCs), which effectively represent the short-term frequency characteristics of audio, emulating the way human ears perceive sound. The audio signals are preprocessed to remove noise, standardized in duration, and passed through MFCC extraction algorithms to create spectrogram-like features. These features form a matrix of coefficients that serve as input to the neural network.

3. CNN MODEL TRAINING MODULE

The **CNN Model Training Module** handles the design, training, and optimization of the Convolutional Neural Network (CNN). The CNN consists of multiple layers, including convolutional layers for spatial feature extraction, pooling layers for downsampling, and fully connected layers for final classification. This module takes the MFCC features as input and trains the CNN to distinguish between gunshot and non-gunshot audio with high accuracy. Techniques like dropout, early stopping, and batch normalization may be employed to improve generalization and avoid overfitting.

4. GUNSHOT CLASSIFICATION MODULE

The **Gunshot Classification Module** is responsible for deploying the trained model to classify new audio inputs. When a user uploads a new audio file, the system extracts MFCC features from it and passes them to the trained CNN. The model then outputs whether the audio contains a gunshot, and optionally, the type of firearm used. This module is key to the real-world application of the system, allowing for live or uploaded audio to be analyzed for potential threats.

5. PERFORMANCE EVALUATION MODULE

The **Performance Evaluation Module** is used to assess how well the CNN performs during training and testing. It calculates various classification metrics such as accuracy, precision, recall, and F1-score, providing a comprehensive view of the model's strengths and weaknesses. Additionally, tools like confusion matrices and loss/accuracy plots over training epochs help in diagnosing issues such as class imbalance or overfitting, guiding further model refinement.

6. WEB INTERFACE MODULE

Lastly, the **Web Interface Module** provides a user-friendly platform through which end-users can interact with the system. Built using HTML, CSS, Bootstrap, and Flask, the interface allows users to upload audio files and view the classification results in real time. The frontend communicates with the backend model hosted on a local or cloud-based WAMP server. This module ensures that the advanced machine learning functionality is accessible to users without requiring technical expertise.

CONCLUSION:

The study successfully demonstrates the effectiveness of using CNNs and MFCCs for accurate and automated gunshot sound identification. Unlike traditional systems, the proposed model eliminates the need for manual feature engineering and expensive hardware. Experimental results indicate high classification accuracy, reduced false positives, and the potential for real-time deployment

in surveillance networks. The ability to distinguish between various firearm types further adds to the utility of the system for public safety and crime prevention initiatives.

RESULT:

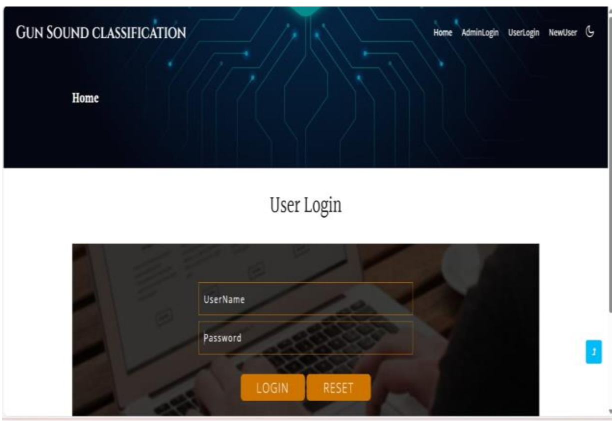

Fig.2 **Fig.3**

FUTURE ENHANCEMENT:

While the current system delivers strong performance, several enhancements are planned to further expand its capabilities:

- **Localization Support**: Integrate GPS or multi-sensor arrays for pinpointing the exact location of the gunshot.
- **Video-Audio Fusion**: Combine audio classification with real-time video analysis for better situational understanding.
- **Edge Deployment**: Optimize the model for deployment on edge devices such as drones, mobile phones, or smart cameras.
- **Noise-Robust Models**: Incorporate noise-invariant training techniques to enhance detection in highly variable environments.
- **Behavioral Threat Detection**: Expand the model to recognize aggressive human behavior or verbal threats using NLP.

These improvements aim to make the system more comprehensive, intelligent, and applicable in a broader range of real-world scenarios.

REFERENCES:

1. Qureshi, Shahzad Ahmad, et al. "Gunshots localization and classification model based on wind noise sensitivity analysis using extreme learning machine." *IEEE Access*, 10 (2022): 87302–87321.
2. Raza, Ali, et al. "Preventing crimes through gunshots recognition using novel feature engineering and meta-learning approach." *IEEE Access*, 11 (2023): 103115–103131.
3. Park, Junwoo, et al. "Enemy Spotted: In-game gun sound dataset for gunshot classification and localization." *IEEE Conference on Games (CoG)*, 2022.
4. Mahmood, Arif, et al. "Detection and localization of firearm carriers in complex scenes for improved safety measures." *IEEE Transactions on Computational Social Systems*, 11.3 (2023): 3900–3910.
5. Vali, Shaik Khaja, et al. "A Deep Learning-Based Efficient Firearms Monitoring Technique for Building Secure Smart Cities." *Material Science*, 23.02 (2024).