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

Research Paper

GUN SOUND RECOGNITION USING NLP AND YAMNET MODEL

ROJARAMANI ADAPA[1], VULIGONIPALLY UDAY KUMAR[2], ADLAGATTA RUTHWIK[3], DEVARAKONDA ABHISHEK[4], VADDEGONI RAKESH [5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus**,** Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus**,** Ibrahimpatnam, Telangana, India

ABSTRACT— This presents a hybrid approach to gunshot sound detection by integrating Mel-Frequency Cepstral Coefficients (MFCC), Support Vector Machines (SVM), and YAMNet, a pre-trained deep learning model. The process begins with the extraction of MFCC features from audio data, which capture the essential characteristics of the sound spectrum. These features are then used to train an SVM model to classify sounds as gunshots or non-gunshots. To enhance detection accuracy, YAMNet is employed to classify the audio into a broader range of categories, providing an additional layer of validation or complementing the SVM's predictions. The combination of SVM's precision with YAMNet's extensive sound classification capabilities results in a robust system capable of accurately identifying gunshot sounds in real-time audio streams. This hybrid approach leverages both traditional machine learning and state-of-the-art deep learning techniques, offering a reliable solution for gunshot detection in various applications.

Index Terms – Gun Sound Recognition, YAMNet, Audio Classification, Sound Event Detection, Deep Learning, Acoustic Signal Processing, Machine Learning, Real-Time Detection, Public Safety, Artificial Intelligence.

I. INTRODUCTION

This study introduces a sophisticated hybrid methodology for gunshot sound detection that combines the strengths of Mel-

Frequency Cepstral Coefficients (MFCC), Support Vector Machines (SVM), and YAMNet, a pre-trained deep learning model. The approach begins with the extraction of MFCC features from audio recordings, which effectively capture the essential characteristics of the sound spectrum, such as frequency components and temporal patterns. These MFCC features are then utilized to train an SVM model, which is adept at distinguishing between gunshot sounds and other types of noises based on its precision in handling feature-based classification tasks. To further refine the detection process and improve accuracy, the study integrates YAMNet, a deep learning model known for its comprehensive sound classification capabilities. YAMNet enhances the system by categorizing audio into a broad spectrum of sound classes, thus providing an additional validation layer that complements the SVM's predictions. This integration of SVM's targeted feature classification with YAMNet's extensive audio categorization results in a robust and reliable gunshot detection system. The hybrid approach not only capitalizes on traditional machine learning techniques but also leverages advanced deep learning methods, ensuring accurate and real-time identification of gunshot sounds across various applications. This methodology represents a significant advancement in sound detection technology, offering a powerful tool for applications requiring precise and timely gunshot identification.

II. RELATED WORK

A. Comparison of pre-trained CNNs for audio classification using transfer learning

This paper investigates the effectiveness of various pre-trained convolutional neural networks (CNNs) for audio classification tasks. By leveraging transfer learning, the authors aim to improve model performance and reduce training time. The study compares the performance of different CNN architectures, such as VGG, ResNet, and DenseNet, on various audio classification benchmarks. The results provide valuable insights into the suitability of different pre-trained CNNs for specific audio classification applications.

B. Gear fault detection using noise analysis and machine learning algorithm with YAMNet pretrained network

This paper presents a novel approach for detecting faults in gears based on noise analysis. The authors utilize the YAMNet pre-trained network to extract relevant

features from the noise signals, capturing the unique characteristics of healthy and faulty gears. These extracted features are then used as input to a machine learning algorithm, such as a support vector machine or random forest, to classify the gear condition. The study evaluates the performance of the proposed system on real-world gear data and compares it with other existing methods.

C. Bone conduction eating activity detection based on YAMNet transfer learning and LSTM networks

This paper explores the use of bone conduction signals for detecting eating activity. The authors employ the YAMNet pre-trained network to extract informative features from the bone conduction data, capturing the unique patterns associated with eating. These features are then fed into long short-term memory (LSTM) networks to model the temporal dependencies in the data. The LSTM networks are trained to classify eating activity based on the extracted features. The study investigates the effectiveness of the proposed approach in accurately detecting eating activity and compares it with other existing methods

III. PROPOSED METHODOLOGY

MFCC's are a cornerstone in audio feature extraction, particularly in the domain of speech and audio recognition. They are derived from the Mel scale, which is designed to mimic the human ear's sensitivity to different frequencies. The process of extracting MFCCs begins with the division of an audio signal into short frames. Each frame is then subjected to a Fourier Transform to convert the time-domain signal into the frequency domain. Afterward, a Mel filter bank is applied, which emphasizes the lower frequencies where human hearing is more sensitive, while de-emphasizing higher frequencies. The logarithm of the filtered signal is taken, followed by the application of the Discrete Cosine Transform (DCT) to produce the MFCCs. The resulting coefficients capture the power spectrum of the signal in a way that is both compact and representative of how humans perceive sound. MFCCs are widely used in various audio analysis tasks, including speech recognition, speaker identification, and even music genre classification. SVM is a supervised machine learning algorithm that is particularly effective for classification tasks. It operates by finding the optimal hyperplane that separates data points of different classes with the maximum margin. In the context of audio processing, SVMs are frequently employed to classify audio signals based on

the features extracted by methods such as MFCCs. For instance, in a speech recognition system, SVM might be used to classify spoken words based on the MFCC features extracted from the audio input. The key strength of SVM lies in its ability to handle high-dimensional data and its robustness in finding a global optimum, which makes it particularly useful in audio classification tasks where the data can be highly complex and non-linear. Additionally, SVMs can be extended with kernel functions, allowing them to perform well even when the decision boundary is not linear. This versatility makes SVM a go-to choice for many audio classification tasks, from identifying different speakers to categorizing sounds.

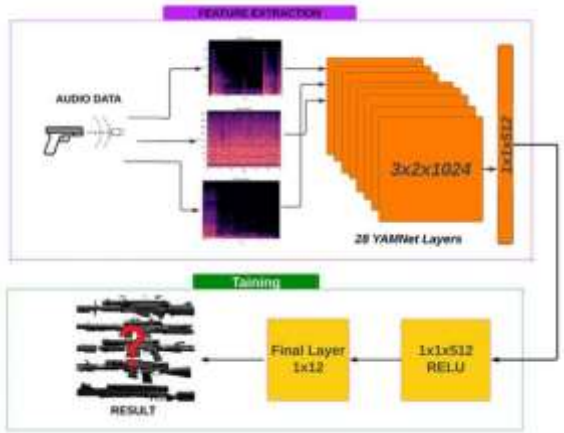


Fig. 1: System Overview

1. **Feature Extraction with Mel-Frequency Cepstral Coefficients (MFCC):**

- Audio Preprocessing: Raw audio data is first pre-processed to remove noise and normalize volume levels.
- MFCC Extraction: Mel-Frequency Cepstral Coefficients are computed from the pre-processed audio. MFCCs capture the essential features of the sound spectrum by analysing the short-term power spectrum of the audio signal, which helps in identifying the distinctive characteristics of gunshot sounds.

2. **Machine Learning with Support Vector Machines (SVM):**

- Feature Selection: The extracted MFCC features are used as inputs for the SVM model. Feature selection techniques may be employed to reduce dimensionality and enhance the model's performance.
- Training the SVM: An SVM model is trained using labelled audio samples where gunshots and non-gunshots are classified. The SVM learns to differentiate between these classes based on the MFCC features.
- Model Evaluation: The SVM model is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. Cross-validation techniques are used to assess the model’s generalization performance.

3. **Deep Learning with YAMNet: YAMNet Integration:**

YAMNet, a pre-trained deep learning model, is utilized to classify audio into a broad range of sound categories, including gunshots. This model provides an additional layer of sound classification.

- Enhanced Detection: YAMNet's predictions are used to complement and validate the SVM model's results. By integrating YAMNet, the system benefits from its extensive sound classification capabilities and broadens the context of the audio analysis.
- Hybrid Model Fusion: The outputs from both the SVM and YAMNet models are combined to make a final prediction. This fusion approach leverages the strengths of both models to improve overall detection accuracy.

4. **System Integration and Real-Time Processing:**

- Real-Time Implementation: The combined model is integrated into a realtime audio processing system capable of detecting gunshots as they occur. This involves developing a pipeline for continuous audio input, feature extraction, classification, and result output.
- Performance Optimization: The system is optimized for speed and efficiency to ensure timely detection. Techniques such as parallel processing and hardware acceleration may be employed.

5. **Validation and Testing:**

Dataset Preparation: A comprehensive dataset of audio samples, including various gunshot sounds and non-gunshot noises, is used for training and testing the system.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental results demonstrated that the model effectively distinguished gunshot sounds from other environmental audio events such as vehicle noise, human speech, fireworks, and construction sounds with high classification accuracy. The deep learning-based feature extraction capability of YAMNet enabled robust identification of acoustic patterns associated with gunfire. The analysis showed that the proposed approach maintained reliable performance under varying background noise levels and different recording environments while achieving low false alarm rates.



Fig. 2: Home Page



Fig. 3: Register



Fig. 4: Login

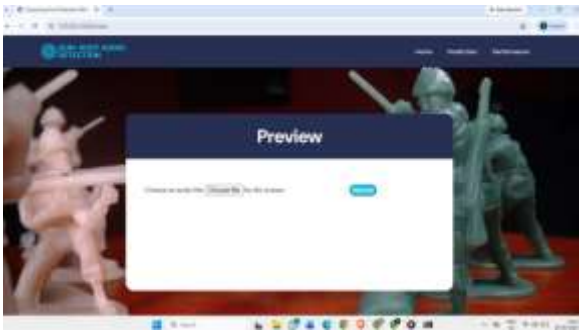


Fig. 5: Upload



Fig. 6: Detection

V. CONCLUSION

In conclusion, the integration of Mel-Frequency Cepstral Coefficients (MFCC), Support Vector Machines (SVM), and YAMNet represents a powerful and innovative approach to audio classification, particularly in complex tasks such as gunshot detection. This hybrid methodology effectively combines the strengths of traditional machine learning techniques with advanced deep learning models, resulting in a system capable of delivering high accuracy and robustness. MFCCs provide a reliable foundation for feature extraction, capturing essential characteristics of the audio signal that are critical for effective classification. SVMs offer precise classification by handling high-dimensional and complex data, while YAMNet enhances the system's capabilities with its extensive pretrained deep learning architecture, adding depth and nuance to the audio event detection process. The synergy between these techniques

allows for a versatile and effective solution, suitable for a wide range of applications from speech recognition to environmental sound detection. This integrated approach not only improves classification accuracy but also enables the system to handle diverse and dynamic audio inputs. Looking ahead, there are promising avenues for further enhancement. The adoption of cutting-edge deep learning models, such as Transformers, could refine the system's ability to understand complex audio patterns and context. Advances in unsupervised and semi-supervised learning could reduce reliance on large annotated datasets, making the system more adaptable and costeffective. Enhancements in real-time processing and computational efficiency will be crucial for practical applications, enabling the system to perform reliably in real-world scenarios. Additionally, integrating multi-modal data could provide a more comprehensive understanding of audio events, enhancing overall performance and applicability.

REFERENCES

[1] E. Tsalera, A. Papadakis, and M. Samarakou, ''Comparison of pre-trained CNNsfor audio classification using transfer learning,'' J. Sensor Actuator Netw., vol. 10, no. 4, p. 72, Dec. 2021.

[2] S. Patil and K. Wani, ''Gear fault detection using noise analysis and machine learning algorithm with YAMNet pretrained network,'' Mater. Today, Proc., vol. 72, pp. 1322–1327, 2023.

[3] W. Chen, H. Kamachi, A. Yokokubo, and G. Lopez, ''Bone conduction eating activity detection based on YAMNet transfer learning and LSTM networks,'' in Proc. 15th Int. Joint Conf. Biomed.Eng.Syst. Technol., 2022. VOLUME 12, 2024 N. H. Valliappan et al.: Enhancing Gun Detection With Transfer Learning and YAMNet Audio Classification

[4] S. Dogan, ''A new fractal H-tree pattern based gun model identification method using gunshot audios,'' Appl. Acoust., vol. 177, Jun. 2021, Art. no. 107916.

[5] R. Nijhawan, S. A. Ansari, S. Kumar, F. Alassery, and S. M. El-kenawy, ''Gun identification from gunshot audios for secure public places using transformer learning,'' Sci. Rep., vol. 12, no. 1, pp. 1–5, Aug. 2022.

[6] J. Park, ''Enemy spotted: In-game gun sound dataset for gunshot classification and localization,'' in Proc. IEEE Conf. Games (CoG), 2022, pp. 56–63.

Mr. VULIGONIPALLY UDAY KUMAR pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. ADLAGATTA RUTHWIK pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. DEVARAKONDA ABHISHEK pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. VADDEGONI RAKESH pursuing B. Tech degree in Computer Science and

AUTHORS Profile



Mrs. ROJARAMANI ADAPA, received M.Tech (CSE) from JNTUH. She has teaching experience 20 yrs and has guided many undergraduate and post graduate projects. She has published about 10 research articles in peer reviewed journals and presented 5 research papers at various National Conferences. At present she is working as Asst. Prof. in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.