Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**

Enhancing gun detection with transfer learning and YAMnet audio classification

#1 B. AMARNATH REDDY, #2 SHAIK INTHIYAZ

#1 Assistant Professor, #2 M.C.A Scholar

DEPARTMENT OF MASTER OF APPLICATIONS

QIS COLLEGE OF ENGINEERING AND TECHNOLOGY

Vengamukkapalem(V), Ongole, Prakasam dist., Andhra Pradesh- 523272

Abstract: Identifying gun types from gunshot audio is very important in areas like forensics, military intelligence, and defence systems. In this study, we provide an improved deep learning architecture that utilises transfer learning with YAMNet for the extraction of features from gunshot noises. We went beyond traditional dense neural network layers and included more advanced designs including CNN1D, CNN2D, CNN3D, and BiLSTM. These architectures let us extract more complex multi-dimensional features from audio spectrograms and waveform patterns.

CNN2D was the most accurate model for this job out of all the ones that were examined. The ADAM and ADAMAX optimisers were used to train and improve the neural networks, while Sparse Categorical Cross-Entropy was used as the loss function. The ADAMAX optimiser stood out because it had a better classification accuracy of 88%, compared to ADAM, which had an accuracy of 84%. The algorithm was tested on a carefully chosen dataset from Kaggle that included audio samples from nine different guns. The models showed that they could generalise well even if the dataset was small. We looked at performance using important measures including accuracy, recall, F1-score, and confusion matrix analysis. The results confirm the effectiveness of deep learning—particularly CNN2D—in the precise categorisation of firearm types based on gunshot audio data, indicating potential applications for real-time surveillance and security systems.

Index terms *- Gunshot Recognition; Audio Classification; Deep Learning; Transfer Learning; YAMNet; CNN1D; CNN2D; CNN3D; BiLSTM; ADAM Optimizer; ADAMAX Optimizer; Sparse Categorical Cross-Entropy; Spectrogram Analysis; Gunshot Audio Dataset; Feature Extraction; Firearm Identification; Forensics; Defense Applications; Real-Time Surveillance; Flask Interface*

1. INTRODUCTION

Gunshot recognition from audio signals has become a critical area of research in modern security, forensic investigations, and military surveillance. The ability to accurately identify the type of firearm based solely on its acoustic signature can support real-time threat detection, crime scene analysis, and automated alert systems. Traditional methods for gunshot detection often struggle with noisy environments and limited feature extraction capabilities, which reduce classification accuracy.

With the advancement of deep learning and transfer learning techniques, it is now possible to extract meaningful features from complex audio signals. YAMNet, a pre-trained model based on MobileNet, effectively captures a broad range of audio characteristics, making it suitable for gunshot feature extraction. In this study, we enhance gunshot classification performance by integrating YAMNet with advanced neural network architectures such as CNN1D, CNN2D, CNN3D, and BiLSTM. This multi-dimensional approach allows for better modeling of gunshot spectrograms and waveform patterns. By leveraging optimizers like ADAM and ADAMAX, we further refine classification performance, achieving notable accuracy even on limited datasets.

This work aims to contribute a robust and scalable system for firearm identification through sound, improving both the efficiency and reliability of automated gunshot recognition systems.

2. LITERATURE SURVEY

2.1 Identification of bullets fired from air guns using machine and deep learning methods:

https://www.sciencedirect.com/science/article/pii/S0379073823001846

ABSTRACT: Ballistics (the linkage of bullets and cartridge cases to weapons) is a common type of evidence encountered in criminal cases around the world. The interest lies in determining whether two bullets were fired using the same firearm. This paper proposes an automated method to classify bullets from surface topography and Land Engraved Area (LEA) images of the fired pellets using machine and deep learning methods. The curvature of the surface topography was removed using loess fit and features were extracted using Empirical Mode Decomposition (EMD) followed by various entropy measures. The informative features were identified using minimum Redundancy Maximum Relevance (mRMR), finally the classification was performed using Support Vector Machines (SVM), Decision Tree (DT) and Random Forest (RF) classifiers. The results revealed a good predictive performance. In addition, the deep learning model DenseNet121 was used to classify the LEA images. DenseNet121 provided a higher predictive performance than SVM, DT and RF classifiers. Moreover, the Grad-CAM technique was used to visualise the discriminative regions in the LEA images. These results suggest that the proposed deep learning method can be used to expedite the linkage of projectiles to firearms and assist in ballistic examinations. In this work, the bullets that were compared were air pellets fired from both air rifles and a high velocity air pistol. Air guns were used to collect the data because they were more accessible than other firearms and could be used as a proxy, delivering comparable LEAs. The methods developed here can be used as a proof-of-concept and are easily expandable to bullet and cartridge case identification from any weapon.

2.2 Gun identification from gunshot audios for secure public places using transformer learning:

https://www.nature.com/articles/s41598-022-17497-1

ABSTRACT: Increased mass shootings and terrorist activities severely impact society mentally and physically. Development of real-time and cost-effective automated weapon detection systems increases a sense of safety in public. Most of the previously proposed methods were vision-based. They visually analyze the presence of a gun in a camera frame. This research focuses on gun-type (rifle, handgun, none) detection based on the audio of its shot. Mel-frequency-based audio features have been used. We compared both convolution-based and fully self-attention-based (transformers) architectures. We found transformer architecture generalizes better on audio features. Experimental results using the proposed transformer methodology on audio clips of gunshots show classification accuracy of 93.87%, with training loss and validation loss of 0.2509 and 0.1991, respectively. Based on experiments, we are convinced that our model can effectively be used as both a standalone system and in association with visual gun-detection systems for better security.

2.3 Enemy Spotted: In-game Gun Sound Dataset for Gunshot Classification and Localization:

https://ieeexplore.ieee.org/abstract/document/9893670

ABSTRACT: Recently, deep learning-based methods have drawn huge attention due to their simple yet high performance without domain knowledge in sound classification and localization tasks. However, a lack of gun sounds in existing datasets has been a major obstacle to implementing a support system to spot criminals from their gunshots by leveraging deep learning models. Since the occurrence of gunshot is rare and unpredictable, it is impractical to collect gun sounds in the real world. As an alternative, gun sounds can be obtained from an FPS game that is designed to mimic real-world warfare. The recent FPS game offers a realistic environment where we can safely collect gunshot data while simulating even dangerous situations. By exploiting the advantage of the game environment, we construct a gunshot dataset, namely BGG, for the firearm classification and gunshot localization tasks. The BGG dataset consists of 37 different types of firearms, distances, and directions between the sound source and a receiver. We carefully verify that the in-game gunshot data has sufficient information to identify the location and type of gunshots by training several sound classification and localization baselines on the BGG dataset. Afterward, we demonstrate that the accuracy of real-world firearm classification and localization tasks can be enhanced by utilizing the BGG dataset.

2.4 Independent Channel Residual Convolutional Network for Gunshot Detection:

https://www.proquest.com/openview/e4516bff97a7c15d0e45aab80940ca2d/1?pq-origsite=gscholar&cbl=5444811

ABSTRACT: The main purpose of this work is to propose a robust approach for dangerous sound events detection (e.g. gunshots) to improve recent surveillance systems. Despite the fact that the detection and classification of different sound events has a long history in signal processing, the analysis of environmental sounds is still challenging. The most

recent works aim to prefer the time-frequency 2-D representation of sound as input to feed convolutional neural networks. This paper includes an analysis of known architectures as well as a newly proposed Independent Channel Residual Convolutional Network architecture based on standard residual blocks. Our approach consists of processing three different types of features in the individual channels. The UrbanSound8k and the Free Firearm Sound Library audio datasets are used for training and testing data generation, achieving a 98 % F1 score. The model was also evaluated in the wild using manually annotated movie audio track, achieving a 44 % F1 score, which is not too high but still better than other state-of-the-art techniques.

2.5 Data Collection, Modeling, and Classification for Gunshot and Gunshot-like Audio Events: A Case Study:

https://www.mdpi.com/1424-8220/21/21/7320

ABSTRACT: Distinguishing between a dangerous audio event like a gun firing and other non-life-threatening events, such as a plastic bag bursting, can mean the difference between life and death and, therefore, the necessary and unnecessary deployment of public safety personnel. Sounds generated by plastic bag explosions are often confused with real gunshot sounds, by either humans or computer algorithms. As a case study, the research reported in this paper offers insight into sounds of plastic bag explosions and gunshots. An experimental study in this research reveals that a deep learning-based classification model trained with a popular urban sound dataset containing gunshot sounds cannot distinguish plastic bag pop sounds from gunshot sounds. This study further shows that the same deep learning model, if trained with a dataset containing plastic pop sounds, can effectively detect the non-life-threatening sounds. For this purpose, first, a collection of plastic bag-popping sounds was recorded in different environments with varying parameters, such as plastic bag size and distance from the recording microphones. The audio clips' duration ranged from 400 ms to 600 ms. this collection of data was then used, together with a gunshot sound dataset, to train a classification model based on a convolutional neural network (CNN) to differentiate life-threatening gunshot events from non-life-threatening plastic bag explosion events. A comparison between two feature extraction methods, the Mel-frequency cepstral coefficients (MFCC) and Mel-spectrograms, was also done. Experimental studies conducted in this research show that once the plastic bag pop sounds are injected into model training, the CNN classification model performs well in distinguishing actual gunshot sounds from plastic bag sounds.

3. METHODOLOGY

i) Proposed Work:

The proposed system enhances gunshot classification by integrating YAMNet-based transfer learning with advanced neural network architectures. Initially, YAMNet extracts high-level audio features from gunshot sounds. These features are then passed into a series of multi-dimensional deep learning models—specifically CNN1D, CNN2D, CNN3D, and BiLSTM—for more effective pattern learning from spectrograms and waveform data.

Among these models, CNN2D demonstrated the highest accuracy by capturing detailed spatial and temporal features from 2D spectrograms. The system

replaces traditional dense layers with these advanced models to improve classification performance. Additionally, training is optimized using ADAM and ADAMAX optimizers, with Sparse Categorical Cross-Entropy as the loss function. This extension not only increases accuracy but also enables the system to handle limited and noisy datasets efficiently. A Flask-based interface allows for user-friendly uploading and real-time audio testing, making it practical for surveillance and forensic applications.

ii) System Architecture:

The proposed system architecture begins with a curated gunshot audio dataset, which is passed through YAMNet for robust feature extraction from the audio signals. The extracted features undergo pre-processing steps such as visualization and data shuffling to ensure consistent training input. These processed features are then split into training and testing sets for model development. Multiple neural network models are trained and evaluated, including a basic neural network with ADAM optimizer, an improved version with ADAMAX optimizer, and an extended model using CNN2D architecture, which achieved superior accuracy. The trained models are assessed using performance metrics like accuracy, precision, recall, and F1-score, which determine the effectiveness of the gun classification task. Finally, based on the best performing model, the system successfully executes gun detection, offering a complete end-to-end pipeline for firearm type recognition using gunshot audio signals.

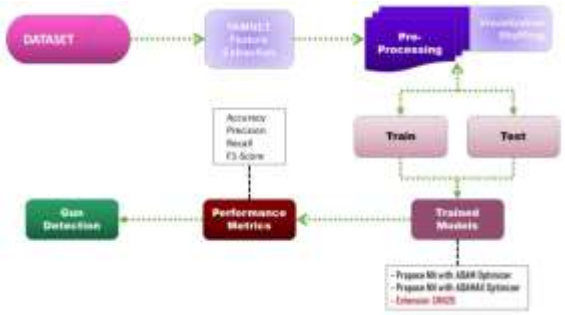

Fig 1 Proposed architecture

iii) Modules:

a. Dataset Collection

- Collect gunshot audio samples from Kaggle consisting of nine different firearm types.
- Ensure data quality and organize into suitable classes for training and testing.

b. Feature Extraction using YAMNet

- Use the pre-trained YAMNet model to extract meaningful features from the raw gunshot audio.
- Capture sound embeddings that represent the acoustic characteristics of each gun type.

c. Data Preprocessing

- Apply visualization and data shuffling to remove bias and enhance training consistency.
- Normalize and structure the extracted features for deep learning input.

d. Model Training

- Train multiple models using different architectures:
    - Basic Neural Network with ADAM Optimizer
    - Neural Network with ADAMAX Optimizer
    - **CNN2D (Extension Model)** for high accuracy

- Use Sparse Categorical Cross-Entropy as the loss function.

e. Model Evaluation

- Evaluate model performance using metrics: **Accuracy, Precision, Recall, and F1-Score**.
- Use confusion matrix analysis for detailed error understanding.

f. Gun Detection Interface

- Develop a Flask-based web interface to allow users to upload audio and get real-time gun type predictions.
- Provide an intuitive and accessible platform for practical use.

iv) Algorithms:

a. YAMNet (Feature Extraction Algorithm)

YAMNet is a pre-trained deep learning model based on MobileNet architecture, designed for audio event classification. It takes raw audio waveforms as input and outputs embedding features representing different sound events. In this project, YAMNet is used to extract relevant features from gunshot audio files, capturing both temporal and frequency domain patterns efficiently.

b. Neural Network with ADAM Optimizer

A basic deep neural network was implemented using the ADAM optimizer, which combines the benefits of AdaGrad and RMSProp. It adjusts the learning rate dynamically for each parameter, making training faster and more stable. This model serves as the initial baseline for firearm classification using the extracted features.

c. Neural Network with ADAMAX Optimizer

This version uses the ADAMAX optimizer, a variant of ADAM based on the infinity norm. It performs better in sparse data conditions and provided higher accuracy than ADAM in this project. It enhances model convergence and helps the network generalize well, especially with limited datasets.

d. CNN2D (Extension Algorithm)

The extension phase introduced CNN2D, a two-dimensional Convolutional Neural Network that processes 2D spectrogram representations of gunshot audio. CNN2D captures both spatial and frequency-based features from the spectrogram, leading to superior accuracy. Among all tested models, CNN2D delivered the best performance for firearm identification.

e. BiLSTM (Bidirectional Long Short-Term Memory)

BiLSTM was also explored as an advanced recurrent neural network that can learn dependencies from both past and future audio contexts. Though it handled temporal patterns well, it was slightly less accurate than CNN2D in this specific task. Still, it showed potential for scenarios requiring long-range audio understanding.

4. EXPERIMENTAL RESULTS

Accuracy: A test's accuracy is determined by its capacity to distinguish between healthy and ill cases. To gauge the accuracy of the test, find the percentage of examined instances that had true positives and true negatives. According to the computations:

$$\text{Accuracy} = TP + TN / (TP + TN + FP + FN)$$

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: Precision is the number of affirmative cases or the classification's accuracy rate. The following formula is applied to assess accuracy:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: A model's ability to recognise every instance of a pertinent machine learning class is measured by its recall. The ratio of accurately predicted positive observations to the total number of positives indicates how well a model can identify class instances.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: Mean Average Precision is one ranking quality metric (MAP). It considers the number of relevant recommendations and their position on the list. MAP at K is calculated as the arithmetic mean of the Average Precision (AP) at K for each user or query.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

F1-Score: An accurate machine learning model is indicated by a high F1 score. combining precision and recall to increase model correctness. The accuracy statistic quantifies the frequency with which a model correctly predicts a dataset.

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

Fig 2 Admin Login

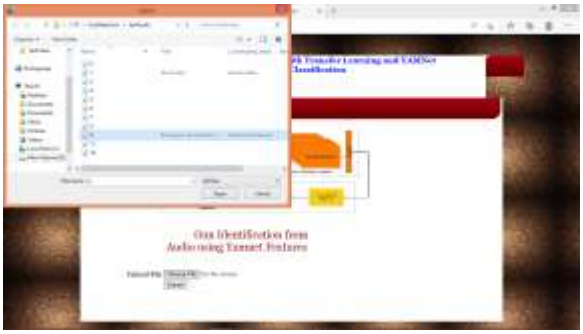

Fig 3 upload image

Fig 4. Results

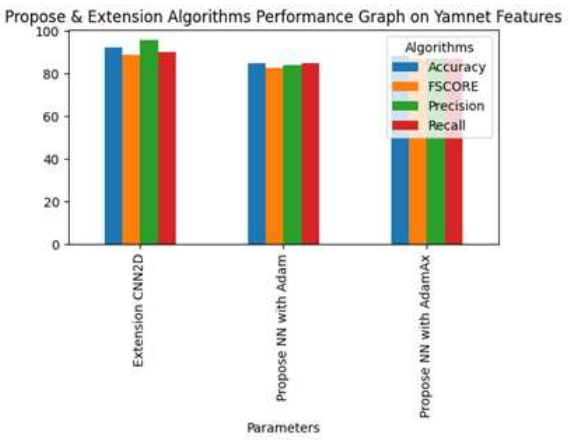

Fig 5 accuracy graph

5. CONCLUSION

In conclusion, the proposed system uses advanced deep learning algorithms to accurately sort weapons by the noises they make when they fire. The system uses YAMNet to get features from audio data and then sends those information to a neural network, where it uses several techniques to improve performance. The ADAM optimiser got 84% accuracy, while the ADAMAX optimiser did better, with 88% accuracy. This shows that it works well for training neural networks for this purpose. The system uses Sparse Categorical Cross-Entropy as a loss function to make learning easier. The model's performance evaluation shows encouraging results, even if the dataset is small and only comprises nine different gun noises from the Kaggle gunshot audio collection. To measure how accurate the classification was, key assessment criteria including precision, recall, and F1 score were used. This showed how neural networks may be used to identify firearms. The ADAMAX optimiser produced the most accurate results, which shows that it is the best choice in this case. This study shows that deep learning architectures may be used to classify sounds, especially in forensics and defence applications. This opens the door for more accurate and efficient gun detection systems that use audio analysis.

6. FUTURE SCOPE

Future work can focus on adding more gunshot sounds to the dataset. This would make the model stronger and better at generalising. Adding sophisticated deep learning architectures like CNNs or recurrent neural networks (RNNs) might make accuracy and feature extraction better. Real-time audio processing might help identify guns right away in dangerous scenarios. Also, looking into transfer learning methods might help improve performance with little data. Working with police and forensic professionals might help make the system better for real-world use by making sure it satisfies the demands of those who work in the field.

REFERENCES

REFERENCES

[1] M. R. K. Mookiah, R. Puch-Solis, and N. Nic Daeid, ‘‘Identification of bullets fired from air guns using machine and deep learning methods,’’ Forensic Sci. Int., vol. 349, Aug. 2023, Art. no. 111734.

[2] 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.

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

[4] J. Bajzik et al., ‘‘Independent channel residual convolutional network for gunshot detection,’’ Int. J.

Adv. Comput. Sci. Appl., vol. 13, no. 4, pp. 950–958, 2022.

[5] R. Baliram Singh, H. Zhuang, and J. K. Pawani, ‘‘Data collection, modeling, and classification for gunshot and gunshot-like audio events: A case study,’’ Sensors, vol. 21, no. 21, p. 7320, Nov. 2021.

[6] 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.

[7] 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.

[8] 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.

[9] J. Bajzik, J. Prinosil, and D. Koniar, ‘‘Gunshot detection using convolutional neural networks,’’ in Proc. 24th Int. Conf. Electron., Lithuania, 2020, pp. 1–5, doi: 10.1109/IEEECONF49502.2020.9141621.

[10] T. Tuncer, S. Dogan, E. Akbal, and E. Aydemir, ‘‘An automated gunshot audio classification method based on finger pattern feature generator and iterative relieff feature selector,’’ Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi, vol. 8, no. 14, pp. 225–243, 2021.

[11] L. G. Martins. (Mar. 2, 2021). Transfer Learning for Audio Data With YAMNet. TensorFlow Blog. [Online]. Available: https://medium.com/analytics-vidhya/understanding-the-mel-spectrogram-fca2afa2ce53

[12] A. Tena, F. Clarià, and F. Solsona, ‘‘Automated detection of COVID-19 cough,’’ Biomed. Signal Process. Control, vol. 71, Jan. 2022, Art. no. 103175, doi: 10.1016/j.bspc.2021.103175.

[13] A. Patel, S. Degadwala, and D. Vyas, ‘‘Lung respiratory audio prediction using transfer learning models,’’ in Proc. 6th Int. Conf. I-SMAC (IoT Social, Mobile, Analytics Cloud) (I-SMAC), Dharan, Nepal, Nov. 2022, pp. 1107–1114.

[14] M. Djeddou and T. Touhami, ‘‘Classification and modeling of acoustic gunshot signatures,’’ Arabian J. Sci. Eng., vol. 38, no. 12, pp. 3399–3406, Dec. 2013, doi: 10.1007/s13369-013-0655-5.

[15] A. K. Sharma, G. Aggarwal, S. Bhardwaj, P. Chakrabarti, T. Chakrabarti, J. H. Abawajy, S. Bhattacharyya, R. Mishra, A. Das, and H. Mahdin, ‘‘Classification of Indian classical music with time-series matching deep learning approach,’’ IEEE Access, vol. 9, pp. 102041–102052, 2021.

[16] N. A. M. Ariff and A. R. Ismail, ‘‘Study of Adam and adamax optimizers on AlexNet architecture for voice biometric authentication system,’’ in Proc. 17th Int. Conf. Ubiquitous Inf. Manage. Commun. (IMCOM), Jan. 2023, pp. 1–4.

[17] H. Ide and T. Kurita, ‘‘Improvement of learning for CNN with ReLU activation by sparse regularization,’’ in Proc. Int. Joint Conf. Neural Netw. (IJCNN), May 2017, pp. 2684–2691.

[18] M. Nicolini, F. Simonetta, and S. Ntalampiras, ‘‘Lightweight audio-based human activity classification using transfer learning,’’ in Proc. 12th

Int. Conf. Pattern Recognit. Appl. Methods. ScitePress, 2023, pp. 783–789.

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

[20] X. Ni, L. Fang, and H. Huttunen, ‘‘Adaptive l2 regularization in person re-identification,’’ in Proc. 25th Int. Conf. Pattern Recognit. (ICPR), Jan. 2021, pp. 9601–9607.

[21] E. Phaisangittisagul, ‘‘An analysis of the regularization between l2 and dropout in single hidden layer neural network,’’ in Proc. 7th Int. Conf. Intell. Syst., Model. Simul. (ISMS), Thailand, Jan. 2016, pp. 174–179, doi: 10.1109/ISMS.2016.14.

[22] M. Wang, S. Lu, D. Zhu, J. Lin, and Z. Wang, ‘‘A high-speed and low-complexity architecture for softmax function in deep learning,’’ in Proc. IEEE Asia–Pacific Conf. Circuits Syst. (APCCAS), Oct. 2018, pp. 223–226.

[23] M. D. Shin, ‘‘Adaptation of pre-trained deep neural networks for sound event detection facilitating smart homecare,’’ J. Abbreviated. [Online]. Available: https://urn.fi/URN:NBN:fi:tuni-202305316375

[24] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, ‘‘MobileNets: Efficient convolutional neural networks for mobile vision applications,’’ 2017, arXiv:1704.04861.

[25] E. Aydemir. (Jun. 22, 2021). Gunshot Audio Dataset. [Dataset]. Kaggle. [Online]. Available: https://www.kaggle.com/datasets/emrahaydemr/gunshot-audio-dataset

[26] R. Lilien and J. Housma. (2019). Gunshot Audio Forensic. Cadre Research. [Online]. Available: http://cadreforensics.com/audio/

[27] (Mar. 6, 2020). Understanding the Mel Spectrogram. Analytics Vidhya. [Online]. Available: https://medium.com/analytics-vidhya/understanding-the-mel-spectrogram-fca2afa2ce53

Author profiles

Mr. B. Amarnath Reddy is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He earned his M.Tech from Vellore Institute of Technology(VIT), Vellore. His research interests include Machine Learning, Programming Languages. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits.