ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/382868138

Real-time Gunshot Detection System Integration To Camera Surveillence System

Preprint · August 2023

DOI: 10.13140/RG.2.2.16098.34249

CITATION
1

READS
647

1 author:

Xinzhang Xiong
Pennsylvania State University
2 PUBLICATIONS **11** CITATIONS

SEE PROFILE

All content following this page was uploaded by Xinzhang Xiong on 05 August 2024.

The user has requested enhancement of the downloaded file.

Real-time Gunshot Detection System Integration To Camera Surveillence System

Xinzhang Xiong
Pennsylvania State University
State College, USA
xpx5059@psu.edu

***Abstract*—The mounting challenge of gun violence necessitates proactive measures, prompting this investigation into deploying a real-time, autonomous gunshot detection system. We envision this system to seamlessly integrate with preexisting infrastructure, such as surveillance cameras, extending their coverage beyond visual cues to incorporate audio signals. This not only circumvents the inherent blind spots or 'dead zones' of cameras but also broadens the spectrum of real-time monitoring and detection. At the heart of this system is a machine-learning model designed for precise audio classification, with a specific focus on gunshots. The model operates efficiently, requiring minimal electrical and computational power, rendering it suitable for integration with existing, resource-limited surveillance systems. We conduct an in-depth exploration of the equilibrium between power efficiency and system performance, concentrating on the system's ability to detect gunshots amid diverse ambient sounds accurately. Moreover, the reliability of the system under real-world operational conditions is thoroughly evaluated. We aim to demonstrate the potential for enhancing current surveillance systems with autonomous, power-efficient, and accurate gunshot detection capabilities. By doing so, this research contributes to the collective endeavor of fostering safer communities.**

Keywords*—Gunshot Detection, Machine Learning, Real-time Audio Processing, Power-Efficient Systems, Camera Surveillance Integration

I. Introduction

The alarming increase in gun violence poses a grave threat to public safety. Since 1970, there have been 1,315 school shootings in the U.S. [1]. In addition, in 2018, there was 97 recorded witnessing shooting incident at k-12 schools in the U.S. [2]. Furthermore, 2022 alone reported an estimated 20,138 firearm deaths, excluding suicides [3]. As substantiated by these statistics, the stark reality of escalating gun violence necessitates more effective detection and alarming system for minimizing the loss of future events. Next, we will discuss the accomplishments and limitations of previous work in this area.

A. Past work

Over the years, a plethora of techniques have been proposed and evaluated for gunshot detection and classification, demonstrating the significance and ongoing interest in this field.

The potential of deep learning architectures in acoustic gunshot detection is established by [4], who proposed a Convolutional Neural Network (CNN) tailored specifically for gunshot classification. Concurrently, [5] developed an innovative solution with an impulsive gunshot recognition method based on energy calculation. This groundbreaking approach, impervious to ambient noise and devoid of any requirement for limits or adaptation, marked a significant leap forward in gunshot detection methodologies.

Building upon the basis of deep learning, [6] ventured beyond simple detection, focusing on discerning the type of firearm (rifle, handgun, or none) based solely on the acoustic signature of the discharge. Their solution employed Mel-frequency-based audio features within a self-attention-based (transformers) architecture, demonstrating the possibilities of using advanced feature extraction and attention mechanisms in this field.

Taking this a step further, [7] introduced a fusion of convolution and recurrent neural networks by proposing a convolution-GRU-based gunshot detection and classification system, demonstrating an average classification accuracy exceeding 80

The applicability of gunshot detection systems extends beyond the realm of human security, as explored by [8]. They employed deep learning for gunshot detection in wildlife conservation areas. This unconventional application of the technology, however, highlighted the susceptibility of these systems to adverse weather conditions like rain and thunderstorms.

Regarding hardware deployment of these models, [9] reported success in training and deploying a CNN model with an impressive accuracy of above 99% on a single Raspberry Pi microcomputer. Similarly, Grane et al. [10] demonstrated the feasibility of deploying a small CNN for real-time gunshot detection on portable cameras, stressing that effective detection does not necessitate extensive memory, considerable battery consumption, or robust CPU power.

These seminal contributions illuminate the path towards developing effective, power-efficient, and robust gunshot detection systems. They reflect the versatility of approaches and applications in this field. However, despite these advancements, considerable challenges persist in real-world deployment and performance under diverse conditions, creating an exigency for further research that this study endeavors to address.

1) Current acoustic Gunshot Detection system: In examining past work on gunshot detection systems, one point of

consensus across the literature is the prohibitive cost of these systems. Acoustic gunshot detection systems are traditionally expensive to implement, and the cost-effectiveness of these systems has been brought into question, especially in settings such as schools where gunshot incidents are rare. According to [11], gunshot events in schools are so rare that the systems are not worth the price tag.

One of the most substantial drawbacks of current acoustic gunshot detection systems is their high cost of implementation. According to [12], the cost of these systems averages between $65,000 and $95,000 per mile of coverage annually. This hefty price tag is further amplified by the necessary investment in infrastructure and maintenance, which includes installation, calibration of the system, and regular updates to ensure optimal performance. Figure.1 further illustrates the cost of traditional ways of conducting gunshot detections.

These cost concerns are not the only challenges associated with traditional gunshot detection systems. Many systems face limitations in terms of coverage area, potential dead zones, and issues of false positives or negatives. Also, power consumption and the need for manual monitoring add further constraints.

Another significant concern with current systems is their tendency to generate false positives, mistaking other loud noises, such as firecrackers or car backfires, for gunshots. According to /citeSPOTFALSEALARM, ShotSpotter false alarms send police on numerous trips (in Chicago, more than 60 times a day) into communities for no reason and on high alert expecting to potentially confront a dangerous situation. This not only compromises the reliability of the system but also could potentially strain the resources of emergency responders.

These considerations underscore the necessity for an efficient, cost-effective, and reliable solution. Our research aims to address these concerns by proposing an autonomous, portable, and power-efficient gunshot detection system, leveraging machine learning techniques for accurate audio classification and integrating with existing surveillance infrastructure for improved detection and response times. In the following sections, we will delve deeper into our proposed system and its potential advantages over current gunshot detection technologies.

	During or After First GUNSHOT		After First GUNSHOT	
	Concealed Weapons Sensors	Metal Detectors	Acoustic Gunshot Detection	Security Officer
Estimated Annual Cost Per Detection Point	$100,000	$25,000-100,000	$75,000	$60,000-80,000 (Annual Salary)
Average Annual Cost per Building	$435,730	$217,865	$490,196	

Fig. 1. Traditional GUNSHOT detection Cost

B. Our work

In this paper, our primary focus is on the deployment of machine learning gunshot detection models, discussed in SectionIII on a variety of portable devices, encompassing Arduino, FPGA, Raspberry Pi, and personal computers, discussed in SectionIV. We confront the critical task of striking a balance between model performance, power consumption, and cost, all while ensuring reliable system functionality. This necessitates careful consideration of trade-offs, given the potential limitations of each hardware platform and the need for an efficient, accurate, and real-time gunshot detection system.

Given the complexity and wide-ranging applications of gunshot detection, it is essential to investigate diverse sound sources, particularly those that may be mistaken for gunshots by the human ear, such as thunderstorms, fireworks, and others. To this end, we have developed a machine learning model for binary classification of gunshot audio against these similar sounding sources. The experimental results, performance insights, and deployment strategies for our model across different hardware platforms will be presented and discussed in detail.

Despite our focus on model development and performance in this paper, we acknowledge the inherent worth of comprehensively exploring and resolving the challenges and considerations associated with the practical realization of such a system. Therefore, we anticipate the continuation of this work in a subsequent publication, which will delve into the intricacies of realizing a fully independent portable system that can be integrated into existing camera surveillance system to enhance the overall performance for the system particularly detecting gunshot events. This paper sets the stage for that discussion, providing a foundation upon which future innovations and improvements can be built. Next section we will discuss some of the challenges encountered while implementing our proposed gunshot detection system integrated to camera surveillance system.

1) Challenges: The development of our proposed gunshot detection system that integrates with existing camera surveillance systems comes with several challenges. In this section, we delve into these issues, including computational constraints and limitations related to machine learning model design and training, unpredictable audio environments, distance coverage limitations of microphones, and the diversity of camera surveillance systems. Each of these challenges must be thoughtfully addressed to create a reliable, cost-effective, and efficient gunshot detection system.

a) Computational Constraint and Limitations: Machine learning models, especially those involving audio data, require substantial computational resources during design and training phases. Deploying these models onto cost-effective devices with limited processing power poses significant challenges. The trade-off between the complexity of the model, its detection accuracy, and the computational resources it requires needs careful consideration.

b) Unpredictable Audio Environment: The performance of our model is significantly influenced by the unpredictable and complex nature of real-world audio environments. Factors such as background noise, echoes, and weather conditions can heavily influence the model's ability to accurately detect gunshots. In particular, the propagation of sound in the physical world is highly dependent on local environmental conditions, including the presence of buildings, terrain, foliage, and other physical structures. These elements can modify the trajectory and characteristics of sound waves, making gunshot detection an arduous task.

Moreover, certain sound events, such as thunderstorms or fireworks, have acoustic signatures similar to gunshots, which can lead to false alarms and significantly impair the system's performance. In specific environments such as subway stations, the ambient noise levels can be significantly high due to frequent train movements, thereby masking the sound of gunshots. This can dramatically reduce the performance of our current model, necessitating further research and adjustments to enhance its resilience against such challenging conditions.

Additionally, the variation in audio signatures between different types of firearms adds yet another layer of complexity to the detection task. Each firearm has a unique acoustic footprint based on its make, model, and ammunition used. Hence, designing a model that can accurately identify a wide range of gunshot sounds while discriminating them from other loud, impulsive noises is indeed a considerable challenge.

c) Distance Coverage Limitation: The efficacy of audio detection systems is intimately tied to the microphone's coverage distance. While expanding the number of microphones could theoretically increase coverage, this approach also escalates the complexity and costs of the system, potentially making it less viable for widespread implementation.

However, the advent of sophisticated algorithms and filters opens up an intriguing possibility: enabling our system to detect gunshots that may be beyond the human auditory range. Acoustic signals associated with gunshots can propagate over large distances, but they may become too faint for human ears at extended ranges. With the appropriate signal processing techniques, our system may be able to pick up and analyze these faint signals, potentially increasing its effective detection range.

However, this approach introduces additional complexities. Signal processing algorithms capable of detecting faint audio signals are often computationally intensive, which could be a limitation for power-constrained systems. Furthermore, the probability of false alarms may increase when attempting to detect such faint signals, given the ubiquity of background noises in most environments. Therefore, while promising, this approach presents substantial challenges that will need to be carefully addressed in future research and system design.

d) Camera Surveillence Variety: The integration of our proposed system with existing camera surveillance systems presents its own set of challenges. The wide variety of surveillance systems, each with their unique hardware configurations and operating conditions, demands our model to be highly adaptable and compatible. Achieving this while ensuring optimum performance is a key challenge that needs to be addressed.

2) Contribution: This study contributes to the burgeoning field of real-time gunshot detection in several key ways:

- Autonomous and Power-Efficient System: Our work is particularly focused on the development of an autonomous, power-efficient gunshot detection system. This perspective is vital for real-world deployment, particularly in situations where power supply might be constrained.
- Novel Machine Learning Models: We have developed and validated two distinct machine learning models. One model leverages TensorFlow's pre-trained YAMNet for audio classification, and the other is grounded on audio Mel Frequency Cepstral Coefficients (MFCC). Both models demonstrate high accuracy and minimal loss, indicating their effective learning from training data and the ability to generalize well to unseen data.
- Hardware Analysis: This research provides a comprehensive examination of four different hardware platforms for deploying our machine learning models. We explore and analyze the feasibility of each hardware in terms of cost, computational capabilities, power consumption, and other relevant aspects, providing crucial insights for future system deployment.
- Benchmarking Real-World Performance: Beyond model training and validation, we evaluate the models on an independent test set, simulating real-world conditions and thus providing a benchmark for system performance outside controlled environments.
- Bridging Theory and Practice: This research endeavors to bridge the gap between theoretical model development and practical application. We offer a detailed discussion on potential challenges that might arise during system deployment, underscoring the need for balance between performance, cost, and processing power.

By advancing these key areas, our research contributes to the ongoing efforts to mitigate the impacts of gun violence through early detection and response systems.

C. Paper Structure

The paper is organized into six comprehensive and interconnected sections, diligently addressing our research objectives, methodologies, findings, and potential future directions.

SectionI serves as the foundation of our study, providing a comprehensive introduction and laying out the urgent issue of escalating gun violence. We discuss the need for real-time gunshot detection systems as a potent countermeasure and underscore the limitations of existing systems. Our research aims to mitigate these gaps are clearly outlined in this section.

SectionII delves into the design of our proposed system. We discuss the system's architecture and functionality, focusing on its integration with existing surveillance systems. Key components are presented, including using pre-existing

microphones for audio input, the processing pipeline, and the alert mechanisms.

SectionIII elucidates the machine learning model used for gunshot detection. We detail the model's training process, using audio data and highlighting the importance of features like Mel Frequency Cepstral Coefficients (MFCC). The deployment of TensorFlow's YAMNet in our model is also discussed.

SectionIV examines the hardware implementation of our model. We explore various hardware options viable for the model's deployment in a power-independent, autonomous system. This includes microcontrollers, Raspberry Pi, FPGA chips, and cloud-based processing platforms.

SectionV presents the results of our machine learning model training for gunshot detection. The model's performance is evaluated using key metrics such as loss and accuracy throughout the training, validation, and testing stages.

Finally, in SectionVI, we articulate the significance of our research findings and outline the potential for future work. This includes the optimization of the model setup for different hardware configurations and expanding the model's capabilities. The potential societal impact of our work is emphasized, marking a significant advancement in the field of gunshot detection systems.

II. System Design

Our proposed gunshot detection system incorporates several critical components working cohesively to ensure real-time and reliable gunshot detection. The central idea behind this design is the seamless integration of the system into pre-existing surveillance camera infrastructure, effectively augmenting their capability with real-time audio detection functionality. This approach not only enhances the efficacy of current surveillance systems but also addresses cost and deployment constraints typically associated with independent gunshot detection systems. In this section, we provide an overview of our system's design, elucidating the functionality of each component, their interactions, and the overall system architecture.

A. System Overview

The system's primary function is to detect gunshots in real-time using audio captured by pre-existing microphones in surveillance cameras. The system processes this audio data, applies a machine-learning model to identify gunshots, and triggers a response sequence upon detection. This system design aims to be low-cost, low-power, and easily integrated into existing surveillance infrastructure.

B. System Components

Microphone Inputs: The system leverages pre-existing microphone infrastructure within the surveillance camera setup to capture environmental sounds. In scenarios where at least three microphones are available and their relative positions can be fed into the algorithm, the system is capable of achieving sound source localization, adding another layer of precision to the detection mechanism.

Audio Processing and Feature Extraction: Upon receiving the raw audio input, the system initiates a sequence of operations for audio processing and feature extraction. Among these processes, Mel Frequency Cepstral Coefficients (MFCC) extraction is a notable one. By applying MFCC along with other operations, we extract valuable audio characteristics or embeddings from the audio signal. These embeddings, which effectively capture the spectral properties of the audio, serve as the input to our pre-trained machine-learning model.

Machine Learning Model: The machine learning model is trained for the specific task of gunshot detection. Leveraging the feature-rich audio embeddings, the model makes real-time decisions regarding the presence or absence of a gunshot sound within the captured audio data. It is crucial to note that the accuracy of our system hinges significantly on the precision of this model.

Alert Mechanism: This component is the final and equally important aspect of the system. Once a gunshot event is detected, the system immediately triggers an alert, allowing for prompt response. The alert mechanism is designed to be flexible, capable of interfacing with various systems - be it directly alerting law enforcement, sending a notification to a control center, or alerting nearby individuals through an audible alarm system. The adaptability of this component underlines the system's potential to be incorporated into a wide array of existing infrastructure setups.

C. System Functionality

The primary function of our system is to detect gunshot events in real-time and swiftly trigger an alert for immediate response. The system begins its operation by capturing environmental audio through the integrated microphones in the surveillance setup. If there are at least three available microphones, the system further estimates the origin of the sound, which aids in pinpointing the location of potential threats. The captured audio signals undergo pre-processing and feature extraction, primarily utilizing Mel Frequency Cepstral Coefficients (MFCC) to generate meaningful audio embeddings. These embeddings capture the unique spectral properties of the audio, which are then fed into our machine learning model. The pre-trained machine learning model, trained specifically for gunshot sound recognition, evaluates these inputs. When it identifies the audio characteristics corresponding to a gunshot, the system triggers an alert mechanism. This alert can be tailored to various systems, such as alerting law enforcement, notifying a control center, or sounding an audible alarm.

D. 2.4 System Architecture

We propose a gunshot detection system that is architected with adaptability and integrability in mind, allowing it to easily fit into existing camera surveillance systems. The system is energy-efficient, with the potential for operation on alternative clean energy sources such as solar or wind power, thereby

highlighting its suitability for extensive deployment. This architecture emphasizes on low power consumption, facilitating a sustainable and environmentally conscious design.

The real-time audio data captured from the surveillance system's pre-existing microphones undergoes a meticulous processing pipeline, as illustrated in Figure2. This pipeline involves passing the audio data through our robust gunshot detection algorithm, the inner workings of which are comprehensively discussed in SectionIII. The algorithm determines if the audio data block contains a gunshot or not.

When the system identifies a gunshot, it triggers the alert mechanism, as visualized in figure2 and figure3 . This mechanism is responsible for activating various response channels to ensure a swift reaction to the potential threat. The interactive and responsive nature of our architecture allows for instantaneous action, providing an enhanced layer of safety in monitored environments.

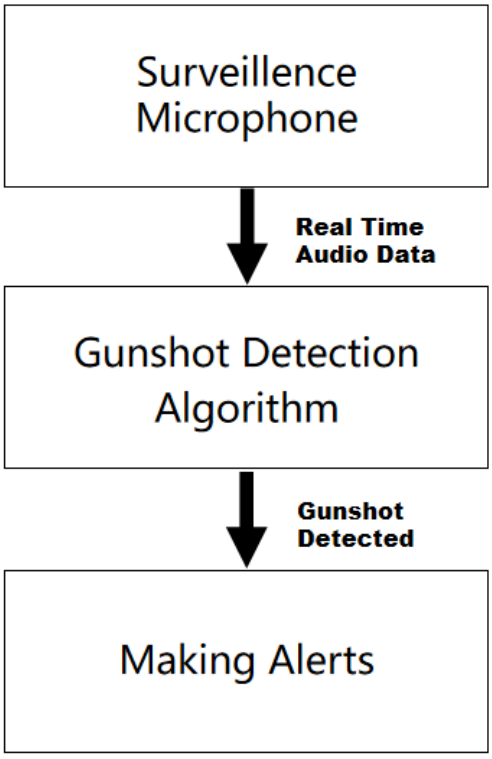

Fig. 2. System Overview

III. Gunshot Detection Machine Learning Model

In the following section, we will introduce an innovative gunshot detection system built using Python, leveraging machine learning to perform binary classification of sound data.

A. Model Overview

The system was trained on a large dataset of 3,200 samples each, consisting of various sounds, including gunshots, barks, door-closing sounds, thunderstorms, and fireworks.

Two distinct models were implemented in the training process. The first model utilized a pre-trained YAMNet model from TensorFlow to extract sound embeddings. This model was further trained using a three-layered neural network, which demonstrated a high level of accuracy, surpassing 99% for training, testing, and validation sets. The second model consisted of a more complex structure with nine layers, including an LSTM layer. This model processed Mel-frequency cepstral coefficients (MFCCs) extracted directly from the audio files and demonstrated similarly high levels of accuracy.

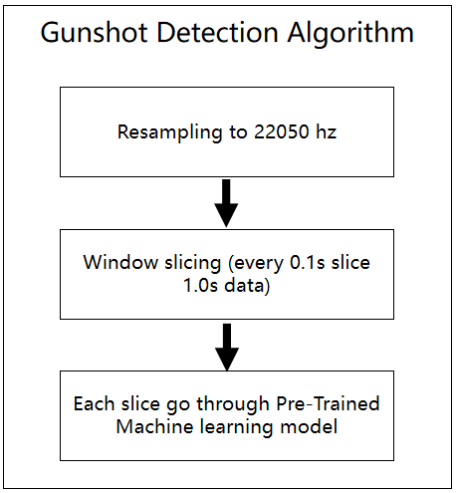

Fig. 3. Algorithm Overview

Gunshot Data Types	Ratio
AK-47	0.085
IMI Desert Eagle	0.118
AK-12	0.115
M16	0.235
M249	0.116
MG-42	0.118
MP5	0.118
Zastave M92	0.096

TABLE I. Gunshot Training Data

The data utilized in this project consists of two main categories: gunshot and non-gunshot (sound similar to gunshot) audio samples. The gunshot data were sourced from the Kaggle Gunshot Audio Dataset, which comprises 851 audio files in total. And the non-gunshot audio primary was downloaded from youtube, containing audio types like a thunderstorm, clapping, and other gunshot-like sounds. The detailed categorical listing of these samples can be found in Table I and II. In the subsequent sections, the utilization of these data will be explained in detail.

B. Training Data

In this section, we will explain the Training data we used to train our Gunshot detection machine-learning model.

1) Gunshot Data: For the Gunshot data group, which includes a wide variety of gunshot sounds from various firearms such as Zestava M92, MP5, MG-42, M249, M16, M4, IMI Desert Eagle, AK-47, and AK12, data were augmented

Non-Gunshot Data Types	Ratio
Thunder	0.008
Snap	0.052
Fireworks	0.286
Drum	0.576
Door	0.026
Clapping	0.012
Bark	0.040

TABLE II. NON-GUNSHOT TRAINING DATA

and resampled into uniform data sets in order to boost the performance of the model.

To prepare the gunshot data for model training, several pre-processing steps were undertaken. A typical audio file magnitude time plot is in Figure. 4 for the gunshot data group. The audio files were determined to first resampled to 1-second, 22050 Hz normalized arrays, with a sliding-window method of 2000hz as shown with orange and green boxes in Table5. The problem is that some files will contain mostly no gunshot, thus making the model trained later, labeling empty files with gunshot or getting confused or unpredictable. To solve this issue, a power filter method was proposed.

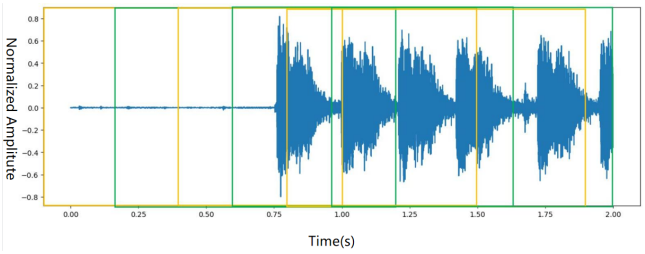

Fig. 4. Typical GUNSHOT Magnitude Plot

The first step is to calculate the total power of each array(all normalized), and the distribution of the total power of the resampled files is shown in Figure. 5. Then we filtered out by setting a power threshold hold to 100 (unit less) to eliminate empty files and other irrelevant audios (this was verified by random sampling by listening to the filtered audio files).

As a result, there are 3210 samples left and available during the next training process, which will be discussed in detail in the methodology section.

C. Non-gunshot data

The non-gunshot data, which served as the negative class for binary classification, were obtained from various sources on YouTube. These samples included thunder, snap, fireworks, drum sounds, door closing, clapping, and barks. The total length of these samples was approximately 3600 seconds. The categorical listing of these non-gunshot audio samples can also be found in Table II. The non-gunshot audio underwent the same pre-processing steps as the gunshot audio, except that no power threshold was set. It was anticipated that the model would classify audio samples with minimal power or

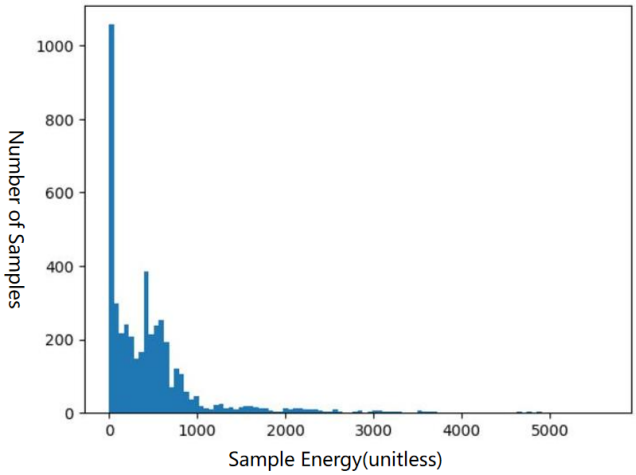

Fig. 5. GUNSHOT audio energy distribution graph

empty audio as non-gunshots, contributing to the model's performance.

As a result, are 7758 samples left and available during the next training process. It will be discussed later that the training data ratio will be balanced and will be discussed in detail in the methodology section.

D. Tensor-Flow Yamnet transferred learning approach

In this section, we will explain the use of TensorFlow's pre-trained YAMNet model for accomplishing the task of binary classification of gunshot and non-gunshot audio files. We implemented a three-layered sequential model to train on the extracted sound embeddings from the YAMNet. The model consisted of an input layer matching the YAMNet's output shape (1024), a dense layer with 512 units using ReLU activation, and an output layer with a neuron count equal to the number of our classes (two, in this case: gunshot and non-gunshot).

After defining the model, we divided our dataset into training, validation, and testing sets. A total of 3200 data samples were used for each group(this compromises the shortage of the gunshot data set), providing a significant base for model training. We used a 60/20/20 split for the dataset, ensuring a robust model performance evaluation.

Once the data was prepared, the model was compiled using the SparseCategoricalCrossentropy loss function from logits, an 'adam' optimizer, and accuracy as the performance metric. We incorporated an EarlyStopping callback to monitor the model's loss during training, halting the training process if no improvement was observed after three epochs and restoring the best weights found during training. This approach allowed us to avoid overfitting and save computational resources. The model was then trained for a total of five epochs. The detailed architecture, data handling, and training process demonstrate the care taken to develop an accurate and efficient gunshot detection system.

E. Traning based on audio MFCC approach

In contrast, another approach we used was based on the audio data's Mel-Frequency Cepstral Coefficients (MFCC). For this method, the same dataset, 3200 samples for each class, was used.

We first converted the raw audio data into MFCC using the Librosa library. This process involved defining a hop length and a number of Fast Fourier Transform (FFT) samples set to 512 and 255, respectively. Each audio data was transformed into an MFCC representation and stored in an array.

To prepare the labels for training, we employed scikit-learn's LabelEncoder to encode the labels. The dataset was then split into training, validation, and testing sets, similar to the previous approach.

The training was performed using a deep learning model with an LSTM (Long Short-Term Memory) layer, which is particularly effective for time-series data such as audio. The model's architecture consisted of an LSTM layer with 128 units, a Flatten layer, two Dense layers with 128 and 64 units, respectively, and three Dropout layers to prevent overfitting. The output layer had 9 units and used softmax activation, as the problem required multiclass classification.

The model was compiled with SparseCategoricalCrossentropy as the loss function, 'adam' as the optimizer, and accuracy as the performance metric. The training was carried out over 50 epochs with a batch size 72. The EarlyStopping callback was also used in this approach, ensuring the model did not overtrain the data. This method illustrates a distinct yet effective way of conducting binary audio classification using MFCC and deep learning.

IV. Hardware Implementation

In developing a gunshot detection system, while the software methodology is premised on machine learning principles, the hardware deployment involves a more nuanced consideration. This section examines four hardware strategies for implementing machine learning models within a power-independent, autonomous system. The primary thrust of this research is centered on exploring systems that manifest a synergistic blend of performance, power autonomy, and operational independence.

In SectionsIV-0a, IV-0b, IV-0c and IV-0c, we examine the use of various hardware platforms, including microcontrollers (with an emphasis on Arduino boards that incorporate microcontrollers into their design), compact computers (exemplified by Raspberry Pi), FPGA chips, and cloud-based processing. Each platform is evaluated based on several crucial parameters, such as cost, computational capability (which directly influences the size of the pre-trained model, thereby impacting the accuracy of the resultant classification), power consumption, and other performance-related factors.

By comparing and contrasting these different approaches, we aim to understand the advantages and trade-offs of each hardware platform, guiding the development of an optimal, efficient, and effective gunshot detection system.

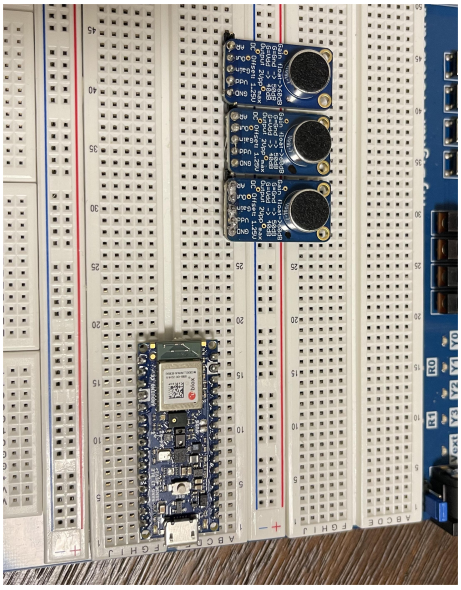

Fig. 6. Microcontroller with microphone

Fig. 7. Resberry Pi

a) Microcontroller and Arduino boards: Microcontrollers and Arduino boards provide an alternative means for deploying real-time gunshot detection systems. These compact and cost-effective devices offer essential computing capabilities while ensuring minimal power consumption, a critical requirement for portable, power-independent systems.

The power consumption of microcontrollers, inherently minimal due to their integrated design, is typically in the milliwatt range. However, this may fluctuate based on the specific model and operational conditions.

From a financial perspective, microcontrollers can range in cost from a nominal amount for basic chips to several tens of dollars for advanced variants. Arduino boards, equipped with built-in microcontrollers and versatile hardware interfaces, generally fall within a price range of $20 to $60, contingent on the specific model.

Regarding physical dimensions, microcontrollers are compact, usually encapsulated within a small chip. Although Arduino boards are somewhat larger due to the additional components, they maintain a modest form factor, often akin to a credit card in size. The diminutive footprint of these

devices is advantageous for their integration into portable or inconspicuous systems.

Despite potentially lacking the processing power of larger computing platforms, microcontrollers, and Arduino boards stand out due to their low power consumption, compact size, and cost-effectiveness. The popularity of the Arduino platform and its strong community support supply abundant resources and pre-existing modules, facilitating the development process.

b) Small computers and Raspberry Pi: Small computers, such as personal computers and Raspberry Pi devices, present another viable platform for implementing real-time gunshot detection systems. These devices offer the benefits of a larger computational capacity than microcontrollers while maintaining a relatively low power consumption and a compact form factor.

Regarding power consumption, personal computers and Raspberry Pi devices are efficient. A typical Raspberry Pi device consumes around 2-3 watts of power under load, while a personal computer's power usage can range widely based on the specific components but is typically much higher.

The cost of these devices is another factor that makes them an appealing choice. A Raspberry Pi can be purchased for as little as $35, while a suitable personal computer may range from a few hundred to a few thousand dollars depending on the specifications. Both options are considerably more affordable than high-end FPGA boards, making them more accessible for widespread deployment.

Regarding form factor, Raspberry Pi devices are notably compact and lightweight, making them suitable for portable or concealed installations. While larger and heavier personal computers can still be conveniently housed in most environments.

The computational power of personal computers and Raspberry Pi devices and their cost-effectiveness and flexibility make them strong contenders for deploying real-time gunshot detection systems. Given their wide adoption and robust community support, they are excellent platforms for developing and deploying such systems.

c) FPGA: Field-Programmable Gate Arrays (FPGAs) offer a unique platform for implementing real-time gunshot detection systems. Due to their parallel processing capabilities and high-performance hardware, FPGAs can efficiently execute machine-learning models for rapid and accurate gunshot detection.

For an FPGA-based system, the device directly processes audio data and generates alerts, effectively reducing data transmission requirements and thus limiting power consumption. The power requirements for FPGA devices can vary substantially based on their specific tasks. Power consumption could range from a few watts to tens of watts for a complex machine-learning task like gunshot detection.

The cost of FPGA devices depends on their capabilities. The cost could start from a few hundred dollars for a basic FPGA board capable of implementing a gunshot detection model. However, higher-end FPGA boards might be required for more sophisticated models or larger datasets, and their cost can extend into the thousands or even tens of thousands of dollars. Despite the higher upfront costs compared to other platforms, FPGAs offer the benefits of real-time processing and flexibility that can be advantageous in certain applications. Regarding deployment, FPGA-based systems can be bulkier and heavier than microcontrollers or personal computers, which might pose challenges for portable or concealed installations. However, their performance advantages may justify these trade-offs in scenarios where real-time response and high accuracy are paramount.

FPGAs have been increasingly recognized for their efficiency and adaptability, making them a promising avenue for the future development of real-time gunshot detection systems. This is particularly true in cloud-based machine learning solutions, where the computational power and customizability of FPGAs are highly beneficial

d) Cloud-based Processing: The advancement of cloud-based machine learning has revolutionized. According to [13], cloud computing can potentially optimize enterprise information systems significantly. Cloud-based machine learning has grown in popularity, it allows for deploying complex machine learning algorithms without the need for extensive hardware infrastructure, thus reducing upfront costs and maintenance efforts.

Applying cloud-based machine learning to audio data analysis, such as in the proposed gunshot detection system, opens up new possibilities for real-time, efficient, and accurate processing. The system's ability to transmit real-time audio data to a data center for analysis and then send back results to the base station for alerts exemplifies the potential of cloud-based machine learning. This section will discuss cloud-based gunshot detection systems regarding cost, power consumption, and data privacy and will address effective real-world deployment in this section.

As per the study by Hamidu et al. [14], a proposed gunshot detection system discussed transmitting real-time audio data to a data center, analyzing and sending back results to the base station for alerts.

Regarding accuracy, this particular framework could yield the highest detection rate(limited by the quality of audio sent to the cloud). From the perspective of keeping the base system as an independent component of the whole system, as for power consumption of the base station, factors such as the quality of the transmitted audio and the base station's components (including the microphone, transmission equipment, and alert receiver) contribute to power usage ranging from a few milliwatts to a few watts. In terms of cost, this model can be comparatively more expensive in the long run. It demands an initial investment in the base station, ongoing carrier charges, data center leasing fees, or the costs of setting up and maintaining a personal server.

V. RESULTS

A. Tensor-Flow Yamnet Transferred Learning Model Result

In the results section, for the performance of the transfer learning model that utilized TensorFlow's pre-trained YAMNet

for audio classification, throughout its training phase, exhibited a commendable performance with a training loss of 0.0114 and accuracy of 99.79

The efficacy of the model was not only restricted to the training and validation phases. When tested on an independent test set, the model continued to perform remarkably well, yielding a loss of 0.0490 and an accuracy of 98.75

To gain further insight into the model's performance, we evaluated its predictions using a confusion matrix as shown in figure8. Out of 1280 samples in the test set (640 from each class), the model correctly classified 641 as gunshots and 631 as non-gunshots. It mistakenly classified 5 non-gunshot sounds as gunshots and 3 gunshot sound as non-gunshots. This minimal misclassification underscores the model's high accuracy and predictive power, making it a promising candidate for real-world deployment in gunshot detection systems.

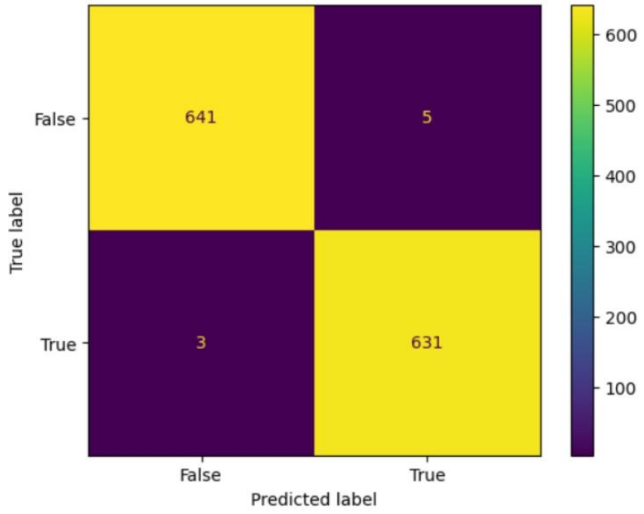

Fig. 8. Model Confusion Matrix

B. Training based on audio MFCC Model Result

For the results obtained from the audio MFCC-based model, during the training phase, this model exhibited superior performance, with a training loss of 0.0068 and an accuracy of 99.84%. In the validation phase, the model continued to show a robust performance, yielding a loss of 0.2039 and an accuracy of 96.56%. These results highlight the model's capacity to learn effectively from the training data while limiting overfitting.

The strength of the model was further demonstrated when it was evaluated using an independent test set. The model returned a loss of 0.1676 and an accuracy of 96.95% on the test data. This high accuracy attests to the model's ability to generalize well to new, unseen data.

Additionally, we evaluated the model's performance using a confusion matrix. Of the 1280 samples in the test set (640 from each class), the model accurately classified 623 as gunshots and 618 as non-gunshots. It incorrectly classified 23 non-gunshot sounds as gunshots and 16 gunshot sound as non-gunshots. While the misclassification rate is slightly higher than the previous model, this MFCC-based model still exhibits high accuracy and predictive power. These results substantiate its potential for effective deployment in real-world gunshot detection systems.

C. Comparative Analysis and Insights

Based on the foregoing analysis of TensorFlow's YAMNet transfer learning model and the audio MFCC-based model, it is evident that gunshot sound classification can be achieved effectively irrespective of hardware limitations. Both models exhibited a high level of proficiency in differentiating gunshot sounds from other audio frequencies that are closely similar. Thus, it's manifest that the critical challenge doesn't stem from the models' classification abilities, but rather from their optimization for efficient real-world deployment.

The primary focus, therefore, transitions to the reduction of the model's size while preserving its high performance. This is a crucial consideration when contemplating the application of these models on smaller, power-efficient chips, which is essential for the realization of fully autonomous and cost-effective systems.

Additionally, although both models demonstrated high accuracy rates, the transfer learning model leveraging YAMNet demonstrated superior performance during the validation and test phases. However, this model's potential demand for more computational resources must be weighed thoughtfully, particularly when aiming for deployment on devices with restricted processing capabilities.

In conclusion, the analysis and insights derived from the study affirm that while both models can indeed be optimized for deployment in real-time gunshot detection systems, there are considerable trade-offs between model size, computational resource requirements, power consumption, and the preservation of high-performance accuracy. These trade-offs are pivotal determinants in deciding the most effective and feasible strategies for model deployment. Future research should, therefore, focus on enhancing these dimensions to strike the optimal balance for the development of real-time, autonomous, and cost-effective gunshot detection systems.

VI. SIGNIFICANCE & FUTURE WORK

This study signifies a crucial step in the ongoing quest to curtail the repercussions of rampant gun violence. The research has successfully demonstrated the viability of using machine learning to distinguish gunshot audio from similar ambient sounds, thereby making it a crucial asset for real-time threat detection systems.

Our proposed methodology extends beyond mere audio classification. It pioneers a path towards implementing the model on a variety of hardware platforms, such as microcontrollers, small computers like Raspberry Pi, and even FPGA chips. Each of these options poses a unique set of benefits and challenges, balancing factors such as cost, computational ability, and power consumption.

Future work should delve further into optimizing the model setup for each hardware configuration. In particular, the primary objective would be to ensure that the system maintains

a high level of performance while operating under stringent computational and power constraints. Therefore, a thorough understanding of the intricacies associated with the deployment of the model on each hardware platform is essential.

Moreover, expanding the range of the model's capabilities to better recognize different types of firearms and gunfire scenarios could significantly enhance its overall performance and real-world applicability. Other avenues for future exploration may include integrating the detection system with emergency response units or developing a multi-modal approach that combines audio with other types of sensor data for more accurate and efficient detection.

In conclusion, this research underscores the potential of machine learning models in developing efficient, portable, and reliable gunshot detection systems. As technology continues to evolve, so does the opportunity to create robust solutions that could play a pivotal role in creating safer societies.

The data utilized in this project consists of two main categories: gunshot and non-gunshot audio samples. The gunshot data were sourced from the Kaggle Gunshot Audio Dataset, which includes a wide variety of gunshot sounds from various firearms such as zestavaM92, MP5, MG-42, M249, M16, M4, IMI Desert Eagle, AK-47, and AK12. The dataset comprises 851 audio files in total. The detailed categorical listing of these samples can be found in Table I.

To prepare the gunshot data for model training, several pre-processing steps were undertaken. The audio files were first resampled to 1-second, 22050 Hz normalized arrays. By calculating the total power of each array, we were able to establish a power threshold to filter out files not containing gunshot sounds. This threshold was identified manually by listening to the audio corresponding to the selected power thresholds, which confirmed the effectiveness of this approach. In addition, to augment the dataset and increase the robustness of the model, a sliding window of 10000 Hz was used to create additional audio samples, effectively doubling the amount of available training data. This step also ensured that the model could accurately detect gunshot sounds, even when only partial gunshot sounds are present in the real-time audio feed.

The non-gunshot data, which served as the negative class for binary classification, were obtained from various sources on YouTube. These samples included thunder, snap, fireworks, drum sounds, door closing, clapping, and barks. The total length of these samples was approximately 3600 seconds. The categorical listing of these non-gunshot audio samples can also be found in Table II. The non-gunshot audio underwent the same pre-processing steps as the gunshot audio, except that no power threshold was set. It was anticipated that the model would classify audio samples with minimal power or empty audio as non-gunshots, contributing to the accuracy of real-world application.

REFERENCES

[1] T. Gibes. Shooting violence in the usa. https://www.defendourchildren.org/.

[2] P. C. Marc A. Zimmerman and R. Cunningham. The facts on children and teens killed by guns. https://www.thetrace.org/2019/08/children-teens-gun-deaths-data/.

[3] C. Brownlee. Gun violence in 2022, by the numbers. https://www.thetrace.org/2022/12/gun-violence-deaths-statistics-america/.

[4] F. Demir, D. A. Abdullah, and A. Sengur, "A new deep cnn model for environmental sound classification," *IEEE Access*, vol. 8, pp. 66 529–66 537, 2020.

[5] Y. Arslan, "Impulsive sound detection by a novel energy formula and its usage for gunshot recognition. arxiv 2017," *arXiv preprint arXiv:1706.08759*.

[6] M. A. Ghazani, A. Hashem-ol Hosseini, and M. D. Emami, "A comprehensive analysis of a laboratory scale counter flow wet cooling tower using the first and the second laws of thermodynamics," *Applied Thermal Engineering*, vol. 125, pp. 1389–1401, 2017.

[7] T. Aggarwal, N. Sharma, and N. Aggarwal, "Gunshot detection and classification using a convolution-gru based approach," in *Proceedings of Emerging Trends and Technologies on Intelligent Systems: ETTIS 2022*. Springer, 2022, pp. 95–107.

[8] D. Danaei, "Gunshot detection in wildlife using deep learning," 2021.

[9] A. Morehead, L. Ogden, G. Magee, R. Hosler, B. White, and G. Mohler, "Low cost gunshot detection using deep learning on the raspberry pi," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 3038–3044.

[10] E. Grane and L. Bokelund, "Gunshot detection from audio streams in portable devices," 2022.

[11] R. Berlin. Benefits and costs of gunshot detection systems in schools. https://www.facilitiesnet.com/emergencypreparedness/tip/Benefits-and-Costs-of-Gunshot-Detection-Systems-in-Schools–39943.

[12] T. Sulzer. Gunshot detection systems vs visual ai gun detection systems. https://zeroeyes.com/gunshot-detection-systems-vs-visual-ai-gun-detection-systems/.

[13] L. Ionescu *et al.*, "Big data analytics tools and machine learning algorithms in cloud-based accounting information systems," *Analysis and Metaphysics*, no. 20, pp. 102–115, 2021.

[14] Hamidu, Ibrahim and Suleiman, Muhammad Aliyu and Abdullahi, Ibrahim, "Smart gunshot detection and reporting framework."

View publication stats