Machine learning inspired efficient acoustic gunshot detection and localization system

Muhammad Salman Kabir[a], Junaid Mir[a] ✉, Caleb Rascon[b], Muhammad Laiq Ur Rahman Shahid[a] and Furqan Shaukat[a]

[a] Electrical Engineering Department, University of Engineering and technology Taxila
[b] Computer Science Department, National Autonomous University of Mexico.

✉junaid.mir@uettaxila.edu.pk

ABSTRACT

Gunshot detection and localization is a frontier technology in security systems. With an increasing rate of shootings globally, gunshot events and directional awareness are crucial for the law enforcement agencies for a timely response. This paper presents a real-time computational efficient gunshot detection and localization system. First, the performance of Mel-frequency cepstral coefficients, linear prediction coefficients, Gammatone cepstral coefficients, and spectral centroid as an audio feature for acoustic gunshot detection is thoroughly analyzed. Then, a bagged tree ensemble and support vector machine classifiers are trained and tested on a diverse gunshot database under different SNR settings, using a 10-fold validation technique. The detection accuracy of 97.3% with a sensitivity of 0.978 and a specificity of 0.988 is achieved. The test-train curves corroborate the fitness and generalization of the trained detection model. After the detection, the localization is performed by calculating the arrival time difference using a general cross-correlation phase transform. Finally, the system is implemented on an experimental test-bed for real-time performance evaluation. Field tests indicate the proposed system's effectiveness to detect and localize a gunshot in 0.7-1 seconds.

Keywords: *Event detection, gunshot detection, audio features, audio localization.*

1. INTRODUCTION

Recently, there has been an alarming rise in gun violence and mass shootings worldwide. According to [1], there were more than 19,000 deaths due to gun usage and 611 mass shooting incidences in the US alone in 2020. In such scenarios, the timely detection and intimation of a gunshot to law enforcement agencies and security professionals is pivotal for prompt and accurate response. Further, situational awareness is even more critical for military operations against terrorism and VIP movements to protect them from gunfire threats and attacks of a trained marksman (i.e., a sniper). In such cases, directional localization after detection is crucial to counteract a threat effectively. Therefore, it is

✉ junaid.mir@uettaxila.edu.pk

interesting to have a system to detect gunshot incidences and help localize and neutralize such threats.

A gunshot event can be detected based on the sound of mechanical action of firearms, the muzzle blast, and the projectile shock wave for supersonic bullets. The acoustic signature of the mechanical action of firearms is generated due to the trigger movement, hammer mechanisms, bullet ejection, and reloading. As the sound intensity of the mechanical action is much lower than the muzzle blast and shock wave, it is not considered for the gunshot event detection [2]. The shock wave is generated by the bullets moving at a supersonic speed and has frequencies ranging from 3000 to 7000 Hz. A muzzle blast is due to the explosion of an explosive charge, which the bullet uses to propel out of a gun barrel. Its acoustic signature usually lasts for 3 to 5 milliseconds with frequencies ranging from 300 to 1000 Hz and can be identified from a distance ranging from several hundred meters to a kilometer [2]. Therefore, the muzzle blast is generally considered the primary acoustical evidence for gunshot detection.

A gunshot event can be detected based on an adaptive threshold or a concise parametric representation (i.e., the acoustic feature) of gunshot acoustical evidence. By using the acoustic features, the acoustic gunshot fingerprints are captured uniquely. Furthermore, the acoustic features are more discriminative and reliable than gunshot audio for gunshot detection. Based on the technique, the acoustic features can be learned or hand-crafted. The learned features are automatically extracted from deep learning (DL) techniques such as neural network (NN), convolution neural network (CNN), recurrent neural network (RNN), etc. In contrast, hand-crafted features require expert feature engineering knowledge and are engineered manually for an application.

The state-of-the-art gunshot detection methods are generally based on acoustic features. For example, Mel-frequency cepstral coefficients (MFCC) [3] and linear prediction coefficients (LPC) [4] are ubiquitous in audio event detection-related studies. However, the contribution of these and other features towards acoustic gunshot detection is not well-assessed in the literature. More importantly, the real-time performance evaluation of gunshot detection and localization is seldom presented in the previous related work. Nonetheless, this is crucial as the real-time performance of gunshot detection systems can be largely influenced due to the noisy conditions. More importantly, it is critical to develop a system with low computational complexity to ensure timely gunshot detection and the localization of the target. However, this important aspect is generally ignored in state-of-the-art.

This paper presents a robust and computationally tractable machine learning-based real-time gunshot detection and localization system. Towards ameliorating this, the assessment of the contribution of four different features for detecting the acoustic gunshot for different noise settings is performed using a challenging, diverse database. It is important to note that, to the author's best of knowledge, among the utilized features, the Gammatone cepstral coefficients (GTCC) [5] have not been used for gunshot detection.

Furthermore, a two-dimensional localization scheme with reduced computational complexity is utilized for gunshot localization that employs a reduced number of microphones. Finally, the real-time performance evaluation is presented through a complete hardware-based field-deployable solution based on the proposed gunshot detection and localization schemes.

The rest of the paper is organized as follows: Section II details the related gunshot event detection work; the proposed methodology for detection and localization is presented in Section III; the experimental setup is elaborated in Section IV; results with discussions are presented in Section V, and the paper is concluded with future work in Section VI.

2. RELATED WORK

Several methods have been proposed in the literature for acoustic gunshot detection. These methods generally rely on detecting a gunshot impulsive acoustic signature and can be categorized into threshold-based and acoustic features-based detection approaches. Threshold-based methods calculate an energy value from the current input window and compare it with prior values [6, 7]. A correlation-based template-matching technique is proposed in [8]. A gunshot is detected if the general cross-correlation of the captured signal against a gunshot template exceeds a pre-defined threshold value. A comparison of template-matching and generalized cross-correlation based on Hilbert kernel spaces is presented in [9]. It concludes that generalized cross-correlation provides more satisfactory results than the template matching technique.

The acoustic feature-based gunshot detection methods can be grouped into techniques that rely on hand-crafted features and those based on learned features. Techniques based on hand-crafted features train a classifier (supervised machine learning (ML) model) on pre-selected features using a set of labelled training data to discriminate between a gunshot and environmental acoustic event. The features employed for training purposes generally include spectral features, energy features, LPC, MFCC, or a combination of these [10-14, 40]. The classification of detected sound based on the extracted features is then done by using neural networks (NN) or ML tools such as support vector machine (SVM) [15, 41], hidden Markov model (HMM) [16], or Gaussian mixture model (GMM) [17], etc. The techniques based on learned features employ deep learning (DL) algorithms to learn and extract suitable gunshot acoustic features in a fully automated manner. In [18], VGG-16 and InceptionV3 CNN architectures trained on 2D representations of audio signals images are used for gunshot detection. A multilayer NN is proposed in [19, 20] for acoustic event detection and compared with HMM for the gunshot classification.

The threshold-based gunshot event detection techniques generally rely heavily on a pre-defined threshold and, therefore, tend to underperform in a noisy environment. Consequently, hybrid techniques are proposed that use thresholding in the pre-processing stage for filtering before the feature extraction [21-23].

Table 1: Summary of the state-of-the-art.

Sr. No.	Method	Approach	Weakness and Strengths
1	Mäkinen et al. [6]	Threshold-based	Derivative of the energy envelope function formed by taking the Short-time Fourier transform of the signal is utilized for gunshot detection.
2	Kotus et al. [7]		A gunshot is detected based on the energy of the signal.
3	Samireddy et al. [8]		General cross-correlation against the signal template is performed for gunshot detection.
4	Djeddou et al. [10]	Handcrafted feature-based	Gunshot detection through GMM trained on MFCC and LPC.
5	Hrabina et al. [11]		Detection is done using a NN trained on different combinations of LPC, LPCC and MFCC.
6	Hrabina et al. [12]		NN trained on low-level time domain features are employed for detection.
7	Singh et al. [13]		Gunshot incidence detected through SVM trained on discrete wavelet transform-based features.
8	Sigmund et al. [14]		Detection through single and ensemble NNs trained on MFCC and time-domain features.
9	Bajzik et al. [18]	Automated Learned feature-based	2D CNN trained on the spectrograms for gunshot detection.
10	Papadimitriou et al. [19]		2D CNN trained on Short-time Fourier transform and MFCC spectrograms for event detection.
11	Conka et al. [20]		Event detection through a recurrent neural network
12	Ahmed et al. [21]	Hybrid	SVM trained on LPC in conjunction with template matching measure used for gunshot detection.
13	Hrabina et al. [22]		Gunshot acoustical detection is based on LPC, sub-band spectral energy comparison, and correlation against a template.
14	Rahman et al.[23]		Detection through ML classifiers trained on antilog energy features coupled with initial threshold-based filtering based on signal energy.

In contrast, though highly accurate and powerful, DL-based techniques are generally complex and demand a large training dataset for learning suitable features for event detection, which is usually scarce for gunshot event detection. Moreover, DL-based techniques can also be compromised in real-time event detection due to overfitting the

training data. Therefore, techniques based on hand-crafted features are generally the most reliable and computationally simple to implement. Table I summarizes the related work.
Once the gunshot incidence is detected, localizing the source is usually solved by estimating the time difference of arrival (TDOA) to calculate its direction of arrival (DOA). This can be carried out by the use of the steered response power phase transform (SRP–PHAT) [24] or generalized cross-correlation phase transform (GCC–PHAT) [25].

3. METHODOLOGY

The block diagram of the proposed acoustic gunshot detection and localization system is shown in Fig. 1, and details of each stage are presented in this section.

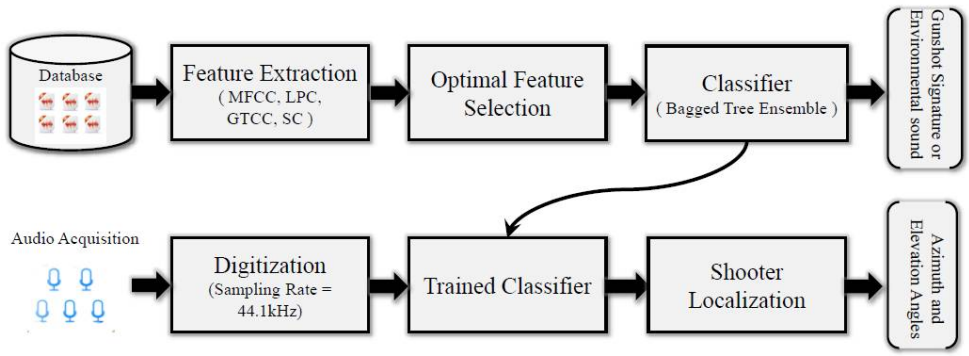

Fig. 1. Proposed system's framework.

A. Training Corpus

Corpora are imperative in developing ML-based detection systems as the trained model's behavior and accuracy depend mostly on the quantity and quality of the training data. To build our training corpus, we used two datasets for each class of interest: gunshot and environment.
Gunshot acoustic samples were acquired from the free firearm library [26], which contains audio samples of AK-47, AR-15, Arisaka, Carl Gustav m/45, Marlin 336, Mosin-Nagant, Norinco SKS, PPSh, Ruger 10/22, Tikka, Winchester and Benelli Nova rifles, machine guns and shotguns gunshots. Each sample has a duration of 3 seconds, sampled at 44.1 kHz. All files were converted to mono WAV format, and then additive white Gaussian noise (AWGN) was added with different signal-to-noise ratio (SNR) values to make the gunshots samples noisy.
The environmental audio samples were obtained from the database named ESC-50 [27]. ESC-50 constitutes environmental sounds like animal voices, vehicles and aircraft sounds, door opening/closing, glass breaking, siren and bell-ringing, etc. [28]. Table II gives the details of the selected samples for the negative environmental sound class. Each sample has a duration of 5 seconds.

The resulting corpus has 1000 audio files, with 250 acoustic samples for the positive class and 750 for the negative class.

B. Feature Extraction

The classification accuracy of any ML-based system depends highly on the features extracted from the input signal. These features should be a concise parametric representation that is more reliable and discriminative for the classification purpose than the actual input signal. To this effect, we propose to use MFCC [3], Gammatone cepstral coefficients (GTCC) [5], LPC [4], and spectral centroid (SC). In Table III, the parameters and mathematical formulas to extract these features are detailed, summarized as follows:

- MFCC models the human auditory system via a set of triangular filters equally spaced on a perceptual pitch scale (aka. Mel scale) [3]. The efficacy of MFCC for audio classification is proven and well-established [29, 30].
- LPC is based on a predictive linear model to imitate the human vocal tract. The LPC coefficients are computed by predicting the frequency and intensity of the residual signal generated by removing the estimated formant effect from the speech signal through inverse filtering [4, 31].

Table 2: Characteristics of the Negative class of Database, i.e., the Environment Sound.

Category	Samples
Animals	20
Water sounds and Natural soundscapes	50
Non-speech and Human sounds	80
Domestic and Interior sounds	200
Urban and Exterior noises	400

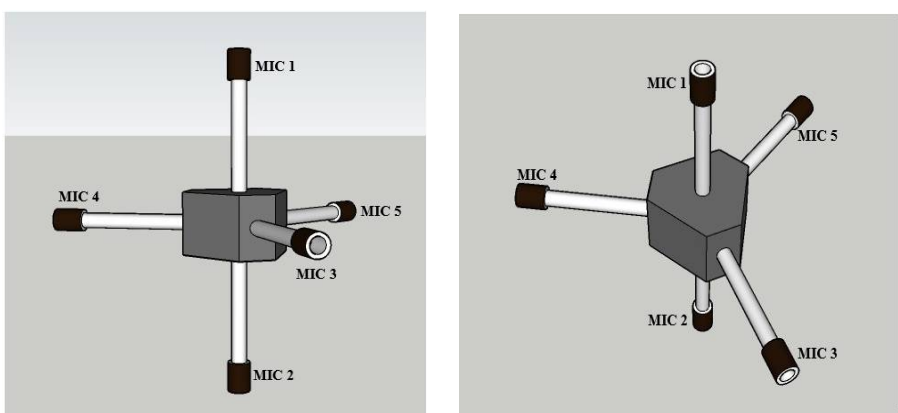

Fig. 2. Microphone array configuration for localization (a) Front view and (b) Top view.

Table 3: Mathematical Expression for Extracting MFCC, GTCC, LPC, and SC Acoustic Features.

Features	Formulas	Description
MFCC	$M(c) = \sum_{f=1}^{F} (\log D(f)) \cos[c\left(p - \frac{1}{2}\right)\frac{\pi}{P}]$	c is the MFCC coefficient index in the range of $[1; C]$; C is the total number of MFCC coefficients; $M(c)$ is the c^{th} MFCC coefficient; $D(f)$ is the mel magnitude of the signal when applying the f^{th} triangular filter; and F is the number of triangular filters.
LPC	$\bar{x}(n) = \sum_{p=1}^{P} a(p)x(n-p)$	$x(n)$ is the n^{th} predicted audio sample; $a(p)$ is the p^{th} LPC coefficient; P is the total number of LPC coefficients; and $x(n)$ is the n^{th} original audio sample.
GTCC	$G(c) = \sqrt{\frac{2}{F}} \sum_{f=1}^{F} (\log D(f) \cos[c\left(f - \frac{1}{2}\right)\frac{\pi}{F}]$	c is the GTCC coefficient index in the range of $[1; C]$; C is the total number of GTCC coefficients; $G(c)$ represents the c^{th} GTCC coefficient; $D(f)$ is the energy of signal in the f^{th} spectral band; and F is the number of GT filters.
SC	$SC = \frac{\sum_f D_m(f).f}{\sum_f D_m(f)}$	D is the input signals power spectrum and f is the frequency index

- GTCC uses a rectangular bandwidth-based Gammatone (GT) filter and is a biologically inspired modification of the MFCC [5]. GT filters model the spectral analysis performed in the cochlea, specifically, the human auditory response.
- The SC estimates the center of mass of the spectrum and is the first order normalized spectral moment.

While both MFCC and LPC have been utilized ubiquitously, GTCC and SC are generally not employed for gunshot detection. Further, these four audio descriptors' individual and combined contribution towards acoustic gunshot detection and classification is not well investigated. The total extracted GTCC and MFCC coefficients are 13 each in the following tests, and the total extracted LPC coefficients are 8.

C. Classifier

For the classification of the acoustic sample, the bagged tree ensemble [32] and the support vector machine (SVM) classifier models are trained and tested on the extracted features. The ensemble method is a supervised ML algorithm in which a new classifier is derived from various base classifiers. For example, an ensemble of bagged tree classifiers combines a decision tree classifier and bagging algorithm. The bagging algorithm (also

called bootstrap aggregation) is one of the most commonly utilized ensemble methods. The derived classifier has a better classification decision performance and robustness due to reduced over-fitting than any constituent base classifiers alone. Moreover, the bagged tree ensemble classifiers are generally more generalized and accurate in out-of-sample testing [33]. SVM works on the principle of margin maximization, where multi-class data is classified into multiple groups by constructing a hyper plan- with maximum possible variance. Both classifiers are trained on the individual and different combinations of the features to assess and investigate their impact on gunshot detection.

D. Localization

The localization is performed by calculating the time difference of arrival of the gunshot acoustic signal, using the GCC-PHAT transform [34]. GCC-PHAT is selected compared to the SRP-PHAT due to its robustness in noisy environments and computational efficiency compared to the SRP-PHAT. In Fig. 2, the five-microphone array employed for gunshot localization is shown. While microphones 3, 4 and 5 are used to estimate the azimuth angle, microphones 1 and 2 are employed to estimate the source's elevation.

After estimating the time difference of arrival of gunshot acoustic, the technique proposed in [35] is used to find out the azimuth angle. First, the direction of arrival (DOA) is estimated by applying a coherence threshold between DOAs found with each microphone pair. Then, after a pairwise estimation of DOA, a redundancy check is performed on the three DOA pairs to find if the three interaural time differences (ITDs) are from a sound source positioned in the same angle sector. The check involves the calculation of the average of the differences between the DOA pairs. Once the DOAs are found to be coherent, the DOA with the minimum absolute ITD value is selected from the DOA set. It is pertinent to mention here that this technique assumes a planar type sound wave, i.e., the sound source is placed in the microphone array's far-field region. Once the azimuth angle is calculated, the binaural sound localization approach uses two microphones to compute the elevation angle.

4. EXPERIMENTAL SETUP

This section details the experimental setup (depicted in Fig. 3), employed to evaluate our real-time deployable gunshot detection and localization system.

A. Microphone Array

The array shown in Fig. 2 employs 5 RTA-M electret microphones [36] which have the following specifications:

- Polar Pattern: Omnidirectional
- Frequency Range: 20 Hz ~ 20 kHz
- Element: Back Electret Condenser
- Impedance: 250 ± 30% (at 1 kHz)
- Sensitivity: -63 ± 3 dB
- Operating Voltage: 9 ~ 52 V DC

- Audio Interface

To digitally capture the five input signals from the microphone array, the open-sourced audio interface 16SoundsUSB from IntRoLab is employed [37]. It is a purpose-built external sound interface with 16 synchronized USB audio inputs. It can capture multi-channel audio synchronously with no time delay present between different channels, which allows the calculation of inter-channel time differences of arrival directly from the captured data. This makes it highly applicable to our case scenario.

B. Processing Unit

A laptop with 8 GB of RAM and a 1 GHz Processor running on Windows 10 is used as a processing unit. The data acquisition toolbox of MATLAB 18b, along with the sound card support, enables the acquisition of the audio signal. The signal is sampled at 44.1 kHz and fed to the trained classifier to detect the acoustic gunshot signatures. Once a positive flag is raised, the acoustic data is sent to the localization module to estimate elevation and azimuth angles.

Fig. 3. Experimental Test-bed.

5. RESULTS AND DISCUSSION

A. Validation Scheme

k- fold cross-validation [38] is used to train and test the classifiers. This validation scheme was selected since it is suitable when the data set size is not considerably large. In a k-fold cross-validation scheme, the data set is randomly split up into k groups (or folds). Then, for each unique group, the group is taken as a test set while the remaining k-1 groups comprise the training set. The model is trained on the training data set and evaluated on

the test data set, thereby resulting in k-1 training of the model. The value of k is 10 in this study.

B. Performance Metrics

To quantify the detection performance of the proposed system, the accuracy, sensitivity, and specificity are computed. In the proposed system's scenario, the sensitivity (also termed as True Positive Rate (TPR) or Recall) is a measure of the proportion of actual gunshot sounds predicted as gunshots. Whereas the specificity (also termed as a True Negative Rate (TNR)) is defined as the proportion of non-gunshot acoustics (i.e., the environmental sounds) that got predicted as non-gunshot sounds. These performance metrics are calculated as,

$$\text{Accuracy}=((TP+TN)/(TP+TN+FP+FN))*100$$

$$\text{Sensitivity or TPR}=TP/(TP+FN)$$

$$\text{Specificity or TNR}=TN/(TN+FP)$$

where TP is True Positive, which means a gunshot sound is classified correctly as a gunshot sound; TN is True Negative, i.e., an environmental sound is classified correctly as an environmental sound; FP is False Positive, meaning an environmental sound is erroneously classified as a gunshot sound; and FN is False Negative, i.e., a gunshot sound is classified incorrectly as an environmental sound.

C. Feature Selection and Simulation Results

To assess the contribution of extracted features towards the acoustic gunshot classification, Table IV presents the 10-fold cross-validation results in terms of accuracy, sensitivity (TPR), and specificity (TNR) for every possible combination of the extracted features. It can be observed that the ensemble bagged tree classifier generally performs better for acoustic gunshot classification problems in terms of performance metrics in comparison to the best performing Quadratic kernel-based SVM. In terms of feature contribution when employed in isolation, a classifier trained on the GTCC feature set has a maximum classification accuracy of 96.3%, as highlighted in Table IV. However, the maximum observed classification accuracy results when features are combined together. As highlighted in Table IV, it can be observed that the best model for gunshot acoustic detection is trained on the combination of LPC, MFCC and GTCC features and has the highest classification accuracy of 97.3% with a sensitivity of 0.978 and specificity of 0.988.

D. Discussion

One of the central problems in ML-based systems is the performance of the classification model in different noise conditions and generalization to other databases in out-of-sample

testing. Table V presents the performance of the bagged tree ensemble classifier under different SNR settings. It can be seen that the LPC based model underperforms in comparison to the GTCC and MFCC based models. For -15 db and -20 db, both GTCC and MFCC based models were similarly effective. However, GTCC based model provides significantly better results than the MFCC based model at -10 db and 0 db. Finally, the model trained on all three features, LPC, MFCC and GTCC, proves to be robust and performs equally well under different noise levels.

Table 4: Classifier Performance: 10-fold Cross-Validation of Bagged Tree Ensemble and Quadratic SVM Classifier for Different Combinations of Features.

Classifiers	**Bagged Tree Ensemble**			**Quadratic SVM**		
Features	**Acc. %**	**TPR**	**TNR**	**Acc. %**	**TPR**	**TNR**
SC	66.5	0.352	0.769	55.7	0.404	0.608
LPC	92.5	0.868	0.944	74.8	0.504	0.997
MFCC	93.9	0.868	0.963	86.0	0.796	0.881
GTCC	***96.3***	***0.892***	***0.987***	87.9	0.864	0.884
SC, LPC	92.4	0.852	0.948	79.2	0.580	0.863
SC, MFCC	95.9	0.900	0.979	89.0	0.876	0.895
SC, GTCC	95.5	0.884	0.979	90.9	0.912	0.908
LPC, MFCC	96.2	0.916	0.977	89.5	0.880	0.900
LPC, GTCC	96.6	0.920	0.981	90.9	0.896	0.913
MFCC, GTCC	96.1	0.864	0.993	88.6	0.860	0.895
SC, LPC, MFCC	95.3	0.888	0.975	91.2	0.884	0.921
SC, LPC, GTCC	96.0	0.896	0.981	92.6	0.924	0.927
SC, MFCC, GTCC	96.5	0.888	0.991	91.1	0.908	0.912
LPC, MFCC, GTCC	**97.3**	**0.978**	**0.988**	90.8	0.888	0.915
All Features	95.9	0.916	0.973	92.1	0.920	0.921

Table 5: Performance of Bagged Tree Ensemble Classifier for Different Noise Levels.

Features	**SNR = -10 *db***			**SNR = -15 *db***			**SNR = -20 *db***		
	Acc. %	**TPR**	**TNR**	**Acc. %**	**TPR**	**TNR**	**Acc. %**	**TPR**	**TNR**
LPC	91.9	0.96	0.91	79.9	0.53	0.89	74.6	0.27	0.91
MFCC	92.2	0.98	0.90	89.4	0.82	0.92	86.3	0.75	0.90
GTCC	95.9	0.98	0.95	89.9	0.81	0.93	86.3	0.74	0.91
LPC, MFCC, GTCC	97.0	0.95	0.98	96.4	0.89	0.99	96.1	0.88	0.99

Table 6: Performance Comparison of the Proposed Method with Related Gunshot detection and Classification Method.

Method	Features	Classification Technique	TPR
2013, Ahmed et al. [21]	LPC	Linear-SVM	97.0%
2018, Hrabina et al. [12]	Scarcely used time domain features	Neural Network	82.2%
2020, Bajzik et al. [18]	Spectrograms	VGG-16	95.8%
2020, Papadimitriou et al. [19]	Spectrograms	2D-CNN	93.0%
2021, Sigmund et al. [14]	MFCC and time domain features	Neural Network	95.0%
2021, Rahman et al. [23]	Antilog energy features	SVM and KNN	93.1%
Proposed	**LPC, MFCC, GTCC**	**Bagged Tree Ensemble**	**97.8%**

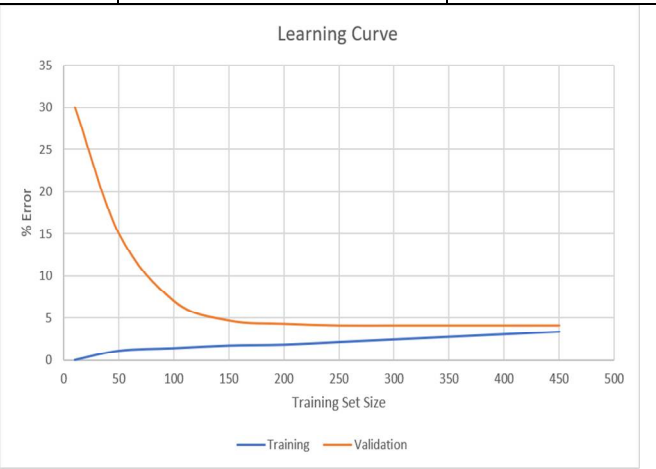

Fig. 4. Training and test error of the best trained classifier.

A well-fitted and generalized model is neither under-fitted nor over-fitted. To find behavior and the generalization of the best-trained model, i.e., the bagged tree ensemble classifier trained on LPC, MFCC and GTCC features, the test and training learning curves are computed (depicted in Fig. 4) to observe the bias-variance trade-off. The low bias, as reflected by the low training error in Fig. 4, indicates the model is well-fitted to the data. Further, the low test error and variance also indicate that the model does not over-fit the data. Therefore, it can be inferred that the trained model is well generalized. The same can also be concluded from the receiver operating characteristics (ROC) curve, where the AUC value of 0.99 is attained, reflecting that the trained model can distinguish between the gunshot and other noises.

Finally, the proposed method performance compared with similar acoustic gunshot detection techniques is summarized in Table VI. However, due to the utilization of the

Table 7: An Overview of Field Tests (Θ = Azimuth, Θ= Elevation).

Firearms	**Observed** (Θ, θ)	**Actual** (Θ, θ)
SIG Sauer P226	1°, -43°	0°, -45°
SIG Sauer P226	62°, -43°	60°, -45°
SIG Sauer P226	-61°, 1°	-60°, 0°
SIG Sauer P226	89°, 1°	90°, 0°
SIG Sauer P226	-92°, 9°	-90°, 10°
SIG Sauer P226	-178°, 9°	180°, 10°
G3	2°, 13°	0°, 15°
G3	57°, 13°	60°, 15°
G3	-59°, 13°	-60°, 15°
G3	92°, 13°	90°, 15°
G3	-87°, 13°	-90°, 15°
G3	-177°, 13°	180°, 15°
MP5	0°, 28°	0°, 25°
MP5	62°, 28°	60°, 25°
MP5	-57°, 28°	-60°, 25°
MP5	88°, 28°	90°, 25°
MP5	-93°, 28°	-90°, 25°
MP5	179°, 28°	180°, 25°
M4 Carbine	1°, -21°	0°, -20°
M4 Carbine	59°, -21°	60°, -20°
M4 Carbine	-61°, 14°	-60°, 15°
M4 Carbine	88°, 14°	90°, 15°
M4 Carbine	-91°, 31°	-90°, 30°
M4 Carbine	-179°, 31°	180°, 30°
AK-47	-3°, -7°	0°, -5°
AK-47	61°, -7°	60°, -5°
AK-47	-59°, 14°	-60°, 15°
AK-47	92°, 14°	90°, 15°
AK-47	-89°, 30°	-90°, 30°
AK-47	178°, 30°	180°, 30°
SSG 69	3°, 59°	0°, 60°
SSG 69	60°, 59°	60°, 60°
SSG 69	-61°, 59°	-60°, 60°
SSG 69	91°, 59°	90°, 60°
SSG 69	-88°, 59°	-90°, 60°
SSG 69	-180°, 59°	180°, 60°

different databases, a one-to-one comparison should be avoided. Further, most of the gunshot acoustic detection-related works have not reported their results quantitatively. Therefore, only a few methods are reported in Table VI for contrast and comparison. It

can be observed that the proposed method performs better than the other techniques for acoustic gunshot detection and classification, as indicated by the TPR (sensitivity) values. The better performance of the proposed method compared to similar hand-crafted feature-based techniques [12, 14, 21, 23] is due to the utilization of the GTCC feature, which establishes the efficacy of the GTCC feature descriptor for gunshot event detection and classification. Interestingly, the DL-based techniques [18, 19] do not outperform the proposed method. One plausible reason for this observation could be the limited availability of gunshot audio databases, which is critical for training DL algorithms.

E. Real-time Localization results in real environment

Field tests were performed to assess the real-time performance of the proposed system and its localization accuracy in a real environment. For this purpose, multiple rounds were fired from SIG Sauer P226, MP5, G3, AK47, M4, SSG 69 and Dragunov firearms in a secured open environment with a distance of 50-1000 meters. While the system successfully detected the fired gunshots, an average error of ±3° is observed in localization, as shown in Table VII. This accuracy is well within the range of acceptable accuracy for real-time sound source localization [39]. Further, the system takes an average of 0.7-1 seconds to detect and localize the gunshots.

6. CONCLUSION

In this paper, efficient real-time gunshot detection and shooter localization system is proposed. The system can efficiently detect and localize a gunshot with an accuracy of 97.3% under ideal conditions. The unavailability of real-time performance evaluation of gunshot detection and localization in previously proposed systems is addressed by evaluating the proposed system on an experimental test-bed and conducting field tests. The well-fitted classifier model reflects that law enforcement agencies can use the proposed system for maintaining law and order. Future works will be aimed toward multi-shot detection and localization of a moving threat.

Acknowledgment: The work and support for real-time gunshot tests were provided by the Heavy Industries Taxila (HIT), Taxila, Punjab, Pakistan.

REFFERENCES

[1] Gun Violence Archive. Available online at https://www.gunviolencearchive.org/ on Dec 02, 2021.
[2] R.C. Maher. Modeling and signal processing of acoustic gunshot recordings, IEEE 12th Digital Signal Processing Workshop & 4th IEEE Signal Processing Education Workshop, 2006, pp. 257–261.
[3] S. Chakroborty, A. Roy and G. Saha. Fusion of a complementary feature set with MFCC for improved closed set text-independent speaker identification, IEEE International Conference on Industrial Technology, 2006, pp. 387–390.
[4] F. Itakura. Line spectrum representation of linear predictor coefficients of speech signals, The Journal of the Acoustical Society of America, Vol. 57(S1), pp. S35–S35, 1975.
[5] X. Valero and F. Alias. Gammatone cepstral coefficients: Biologically inspired features for non-speech audio classification, IEEE Transactions on Multimedia, Vol. 14(6), pp. 1684–1689, 2012.
[6] T. Mäkinen and P. Pertilä. Shooter localization and bullet trajectory, caliber, and speed estimation based on detected firing sounds. Applied acoustics, Vol. 71(10), pp. 902–913, 2010.

[7] J. Kotus, K. Lopatka and A. Czyzewski. Detection and localization of selected acoustic events in acoustic field for smart surveillance applications, Multimedia Tools and Applications, Vol. 68(1), pp. 5–21, 2014.
[8] S.R. Samireddy, J. Carletta and K.-S. Lee. An embeddable algorithm for gunshot detection, IEEE 60th International Midwest Symposium on Circuits and Systems, 2017, pp. 68–71.
[9] J. Van der Merwe and J. Jordaan. Comparison between general cross correlation and a template-matching scheme in the application of acoustic gunshot detection, IEEE Africon, 2013, pp. 1–5.
[10] M. Djeddou and T. Touhami. Classification and modeling of acoustic gunshot signatures, Arabian journal for science and Engineering, Vol. 38(12), pp. 3399–3406, 2013.
[11] M. Hrabina and M. Sigmund. Comparison of feature performance in gunshot detection depending on noise degradation, IEEE 27th International Conference Radioelektronika (RADIOELEKTRONIKA), 2017, pp. 1–4.
[12] M. Hrabina and M. Sigmund. Gunshot recognition using low level features in the time domain, IEEE 28th International Conference Radioelektronika (RADIOELEKTRONIKA), 2018, pp. 1–5.
[13] V. Singh, K.C. Ray and S. Tripathy. Robust gunshot features and its classification using support vector machine for wildlife protection, Electronic Systems and Intelligent Computing, pp. 939-948), 2020.
[14] M. Sigmund and M. Hrabina. Efficient feature set developed for acoustic gunshot detection in open space, Elektronika Ir Elektrotechnika, Vol. 27(4), pp. 62-68, 2021.
[15] Pisner, A. Derek and David M. Schnyer. Support vector machine. Machine Learning. Academic Press, 2020.
[16] C. Viroli and G.J. McLachlan. Deep Gaussian mixture models, Statistics and Computing, Vol. 29(1), pp.43-51, 2019.
[17] L. Rabiner and B. Juang. An introduction to hidden Markov models, IEEE ASSP Magazine, Vol. 3(1), pp. 4-16, 1986.
[18] J. Bajzik, J. Prinosil and D. Koniar. Gunshot detection using Convolutional neural networks, IEEE 24th International Conference Electronics, 2020, pp. 1–5.
[19] I. Papadimitriou, A. Vafeiadis, A. Lalas, K. Votis and D. Tzovaras. Audio-based event detection at different SNR settings using two-dimensional spectrogram magnitude representations, Electronics, Vol. 9(10), pp.1593, 2020.
[20] D. Conka and A. Cizmar. Acoustic events processing with deep Neural network, IEEE 29th International Conference Radioelektronika, 2019, pp. 1–4.
[21] T. Ahmed, M. Uppal and A. Muhammad. Improving efficiency and reliability of gunshot detection systems, IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 513–517.
[22] M. Hrabina and M. Sigmund. Acoustical detection of gunshots, IEEE 25th International Conference Radioelektronika (RADIOELEKTRONIKA), 2015, pp. 150–153.
[23] S.U. Rahman, A. Khan, S. Abbas, F. Alam and N. Rashid. Hybrid system for automatic detection of gunshots in indoor environment, Multimedia Tools and Applications, Vol. 80(3), pp. 1–11, 2021.
[24] A. Sazontov, I. Smirnov and A. Matveev. Source localization in a shallow-water channel with a rough surface, Acoustical Physics, Vol. 61(1), pp. 109–116, 2015.
[25] P. Volgyesi, G. Balogh, A. Nadas, C.B. Nash and A. Ledeczi. Shooter localization and weapon classification with soldier wearable networked sensors, ACM Proceedings of the 5th international conference on Mobile systems, applications and services, 2007, pp. 113–126.
[26] Sound Effects Library. Available online at https://www.airbornesound.com/ on Dec 02, 2021.
[27] K.J. Piczak. ESC: Dataset for environmental sound classification, ACM 23rd international conference on Multimedia, 2015, pp. 1015–1018.
[28] ESC-50 Database. Available Online at https://github.com/karolpiczak/ESC-50/ on Dec 02, 2021.
[29] M.S. Ahmad, J. Mir, M.O. Ullah, M.L.U.R Shahid, and S.M. Adnan. An efficient heart murmur recognition and cardiovascular disorders classification system, Australasian Physical and Engineering Sciences in Medicine, Vol. 42(3), pp. 733-743, 2019.
[30] S. Salman, J. Mir, M.T. Farooq, A.N. Malik, and H. Rizki. Machine learning inspired efficient audio drone detection using acoustic features, IEEE International Bhurban Conference on Applied Sciences & Technology (IBCAST), 2021, pp. 335-339.
[31] S.A. Alim and N.K.A. Rashid. Some commonly used speech feature extraction algorithms, From Natural to Artificial Intelligence-Algorithms and Applications, IntechOpen, 2018.

[32] S. González, S. García, J.S. Del, L. Rokach, and F. Herrera. A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities, Information Fusion, Vol. 64, pp. 205-237, 2020.
[33] Y. Bian and H. Chen. When does diversity help generalization in classification ensembles?, IEEE Transactions on Cybernetics, 2021.
[34] M. Omologo and P. Svaizer. Acoustic event localization using a cross power-spectrum phase based technique, IEEE International Conference on Acoustics, Speech and Signal Processing (ICAASP), 1994, pp. II-273.
[35] C. Rascon, G. Fuentes and I. Meza. Lightweight multi-DOA tracking of mobile speech sources, EURASIP Journal on Audio, Speech, and Music Processing, Vol. 1, pp. 1-16, 2015.
[36] RTA-M: Reference Microphone. Available Online at https://dbxpro.com/en/products/rta-m on Dec 02, 2021.
[37] 16SoundsUSB. Available Online at https://github.com/introlab/16SoundsUSB on Dec 02, 2021.
[38] T. Fushiki/ Estimation of prediction error by using K-fold cross-validation, Statistics and Computing, Vol. 21(2), pp. 137– 146, 2011.
[39] C. Rascon and I. Meza. Localization of sound sources in robotics: A review, Robotics and Autonomous Systems, Vol. 96, pp. 184–210, 2017.
[40] B. Singh, Rajesh, H. Zhuang, and J.K. Pawani. Data collection, modeling, and classification for gunshot and gunshot-like audio events: a case study, Sensors, Vol. 21, pp. 7320, 2021.
[41] B. Tardif, Lo David, and R. Goubran. Gunshot Sound Measurement and Analysis, IEEE Sensors Applications Symposium (SAS), 2021.

Muhammad Salman Kabir is graduated from UET Taxila, Pakistan, in 2019, as an electrical engineer with a specialization in communication. He is currently pursuing a master's degree in computer science from Innopolis University, Russia. His research interests are in signal processing, machine learning and neuroscience.

Junaid Mir is an Assistant Professor in the Department of Electrical Engineering UET Taxila, Pakistan. He received the B.Sc. degree (Hons.) and the M.Sc. degree in Electrical Engineering from the UET Taxila, in 2008 and 2011, respectively. He was awarded a Ph.D. degree from the University of Surrey, U.K. in 2017. His research interest lies in signal, image and video Processing.

Caleb Rascon is an Associate Researcher in the Department of Computer Science within the Institute of Applied Mathematics and Systems Research, National Autonomous University of Mexico. He received his Ph.D. in Electrical and Electronic Engineering from the University of Manchester, UK. His research interest lies in auditory scene analysis, artificial analysis and control and optimization.

Muhammad Laiq Ur Rahman Shahid received the B.Sc. Engg. and M.Sc. Engg. Degrees in Electrical Engineering from the UET Taxila, Pakistan, in 2008 and 2012 respectively. He was awarded Ph.D. degree from Jacobs University Bremen, Germany in 2016 where his research area was medical image processing. He is currently working in the UET Taxila, Pakistan as an Assistant Professor

Furqan Shaukat is currently working as an Associate Professor in the Department of Electrical Engineering UET Taxila. He completed his B.Sc. in Electrical Engineering from the UET Lahore in 2007. He received his M.Sc. and Ph.D. degrees from the UET Taxila in 2011 and 2018, respectively. His research interests include medical image analysis and classification.