IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 August 2023, accepted 12 September 2023, date of publication 18 September 2023,
date of current version 26 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3316695

RESEARCH ARTICLE

Preventing Crimes Through Gunshots Recognition Using Novel Feature Engineering and Meta-Learning Approach

ALI RAZA[1], FURQAN RUSTAM[2], BHARGAV MALLAMPATI[3], PRADEEP GALI[3], AND IMRAN ASHRAF[4]

[1]Institute of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan 64200, Pakistan
[2]School of Computer Science, University College Dublin, Dublin, D04 V1W8 Ireland
[3]Department of Electrical Engineering, University of North Texas, Denton, TX 76203, USA
[4]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, South Korea

Corresponding authors: Furqan Rustam (furqan.rustam@ucdconnect.ie) and Imran Ashraf (ashrafimran@live.com)

This work was supported by University College Dublin, Ireland.

ABSTRACT Gunshot sounds are common in crimes, particularly those involving threats, harassment, or killing. The gunshot sounds in crimes can create fear and panic among victims, often leading to psychological trauma. Gunshot sounds are associated with a significant mortality rate, especially in cases of gun violence. The sound of gunshots can serve as evidence in criminal investigations, allowing law enforcement officials to determine the number of shots fired, the caliber of the gun used, and the distance from which the shots were fired. Efficient gunshot detection is necessary to address the issue of gun violence in society. This study aims to detect gunshot sounds using an efficient approach to prevent crimes. The frequency-time domain spectrum analysis is performed to understand the patterns of signals related to each target class. A novel Discrete Wavelet Transform Random Forest Probabilistic (DWT-RFP) feature engineering approach is proposed, which takes Mel-frequency cepstral coefficients (MFCC) extracted from gunshot sound data as input for feature extraction. A novel meta-learning-based Meta-RF-KN (MRK) is proposed to detect gunshots based on newly created ensemble features with a DWT-RFP approach. For experiments, the gunshot sounds dataset containing 851 audio clips collected from public videos on YouTube from eight kinds of gun models, is used. Advanced machine learning and deep learning techniques are applied in comparison to evaluate the performance of the proposed approach. Extensive experiments show that the proposed MRK approach achieves 99% k-fold accuracy for detecting gunshots and outperforms state-of-the-art approaches. The proposed approach can potentially be used for accurate gunshot detection and to help prevent crimes.

INDEX TERMS Gunshot sound classification, crime prevention, signal processing, MFCC features, meta-learning, feature engineering.

I. INTRODUCTION

The field of audio forensics has seen significant advancements in recent years to identify and analyze gunshot sounds through gunfire recognition techniques [1]. With rising gun-related crimes globally, it is vital to accurately distinguish firearm noises from other sounds within the recorded audio, aiding corporate companies in tackling heightened security challenges [2]. Digital signal processing (DSP), pattern recognition (PR), and machine learning (ML) techniques are utilized in the classification process thus allowing effective distinction among different firearm sounds. This tactic serves as a valuable resource not just limited to forensic experts; it also has potential applications as an anti-cybercrime tool. Cybercriminals tend to simulate firearms through online platforms such as games or social media [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Zia Ur Rahman.

Cybersecurity professionals can use gunfire recognition technology to identify and trace the sound source to prevent future attacks. Furthermore, these technologies can detect gunshot origins in public places and serve as an early warning system for law enforcement agencies [4]. This could help them respond better and quickly to potential threats while ensuring public safety. Ultimately utilizing gunfire recognition technology in audio forensics holds significant potential for preventing and investigating crimes while serving as a critical aspect of corporate security assurance programs globally.

The confluence of mortality rates and cybercrime involving gunshots is a pressing issue in the United States [5]. Although there is limited research on the specific relationship between gun violence and cybercrime, studies have established a correlation between higher rates of gun-related deaths and increased access to firearms [6]. Furthermore, cybercrime can facilitate the spread of illicit firearms via online marketplaces and other channels [7]. As technology advances, it is crucial to account for the potential consequences of cybercrime on public health and safety, particularly in the context of firearm-related mortality rates. Addressing this problem necessitates a multifaceted strategy involving improved online gun sales regulation, stricter enforcement of current gun laws, and targeted interventions to prevent cybercrime and reduce gun violence.

Identifying and locating gunshots in real-time is essential for law enforcement, security, and military operations. Effective gunshot recognition systems have been developed with the latest artificial intelligence techniques and signal processing advancements [8]. These systems implement machine learning algorithms like neural networks to analyze audio signals and distinguish gunshots from similar sounds, like fireworks or car backfires. Furthermore, signal processing techniques including time-frequency analysis are utilized to extract audio signal features and enhance gunshot detection accuracy [9]. Consequently, gunshot recognition systems have the potential to save lives by warning of firearm incidents early and assisting law enforcement in their endeavors to combat gun violence. More research is necessary to enhance the effectiveness and efficiency of these systems and make them more widely accessible. The main contributions of this study for gunshots recognition are as follows

- A novel discrete wavelet transform-random forest probabilistic (DWT-RFP) feature engineering approach is proposed, which utilizes Mel-frequency cepstral coefficients (MFCC) extracted from gunshot sound data as input for feature extraction. The DWT and random forest-based probabilistic features are extracted from MFCC, creating a new feature set to build a machine learning model. In addition, frequency-time domain spectrum analysis is performed on gunshot sound data to understand the patterns of signals related to each target class. Experiments reveal that high-performance scores can be achieved with the proposed DWT-RFP approach.
- A novel meta RF-KN (MRK) model is proposed to detect gunshots from the newly created feature set. The DWT-RFP-based feature set is input to the KN and RF models. The predictions from these models are combined and input to a random forest-based meta-learner for final prediction. This process enhances the performance and provides a highly accurate performance for gunshot detection.
- Seven advanced machine learning and deep learning models are applied for gunshot detection for performance comparison. Applied models are optimized using a hyperparameter tuning approach. The performance of each method is validated using a k-fold cross-validation mechanism.

The remaining research is formatted as Section II comparatively evaluate the related literature for sound classification approaches. Section III describes the study material and methods. The results of applied methods are comparatively discussed in Section IV. The proposed study findings are summarized in Section V.

II. RELATED WORK

The related work section for sound classification techniques with machine learning typically explores prior research in audio classification, as analyzed in Table 1. This section comprehensively reviews existing techniques, datasets, and evaluation metrics used for sound classification. This related work section critically analyzes the current state-of-the-art techniques in sound classification with machine learning and the foundation for the proposed research to address the existing gaps and challenges in this field.

The research article [10] proposes a method for detecting and classifying security sound events using convolutional neural networks (CNN). The system utilizes surveillance camera systems with integrated microphones. The dataset based on spectrogram images of sounds is used to train the applied CNN model. The system achieved a high accuracy of 92% on the training dataset and 90% on the testing dataset. The proposed system's objective is to detect security events with low sound levels and improve the security of the surveillance system.

The study [11] suggests a deep learning method called AReN to identify particular sound events in audio surveillance, such as screams, broken glasses, and gunshots. To represent the audio stream, a gammatonegram image is used, and the 21-layer CNN is fed with parts of this representation. The proposed technique includes problem-driven data augmentation that enhances the training dataset with gammatonegram images obtained from sounds having diverse signal-to-noise ratios. The AReN model yielded a performance accuracy score of 91.43% using the freely accessible SESA dataset.

The authors discussed the use of a CNN for classifying urban sounds in [12]. The study involved training the CNN model with short audio signals and spectrograms from 10 distinct sound categories. The audio was transformed into Mel spectrograms and treated as visual images for CNN. The final model achieved a 91% accuracy rate on the test dataset, which suggests that the method effectively

TABLE 1. Analysis of gunshot sound-related literature.

Ref.	Year	Dataset	Technique	Performance Score
[10]	2021	Spectrogram images of sounds.	Convolutional neural network	Accuracy: 92%
[11]	2020	SESA data based on gammatonegram images.	AReN	Accuracy: 91%
[12]	2021	Short audio signals and spectrograms.	Convolutional neural network	Accuracy: 91%
[13]	2021	UrbanSound8K	Convolutional neural network	Accuracy: 94%
[14]	2022	AudioSet and the NIJ Grant 2016-DN-BX-0183 gunshot dataset.	Teacher network	Accuracy: 95%
[15]	2022	ESC dataset contains 5000 sounds.	k nearest neighbors	Accuracy: 93%
[16]	2022	Data of 597 gunshots and 28,195 background sounds.	Convolutional neural network	Recall: 95%
[17]	2023	UrbanSound8K	Hybrid ensemble classifier	Accuracy: 79%
[18]	2020	ESC-10 sound signals data.	Support Vector Machines	Accuracy: 94%
[19]	2021	UrbanSound8K	Convolutional neural network	Accuracy: 96%

categorizes urban sounds. Similarly, [13] presents a novel approach for classifying environmental sounds, which involves using a one-dimensional CNN along with Bayesian optimization and ensemble learning. The proposed end-to-end model directly extracts feature representations from audio signals through convolutional layers that capture signals and learn relevant filters for classification. Hyper-parameters selection and model evaluation was carried out using Bayesian optimization with cross-validation. To test the model's performance, the UrbanSound8K dataset was used, achieving an accuracy of 94.46% with a reduced number of trainable parameters.

Along the same directions, the study [14] proposed a rapid gunshot-type recognition method based on knowledge distillation to address the need to reduce the size of a gunshot recognition network model and improve real-time detection in urban combat. To achieve this, the authors suggest preprocessing the muzzle blast and shock wave generated by the gunshot and enhancing the dataset's quality through the Log-Mel spectrum. The researchers constructed a teacher network with 10 two-dimensional residual modules and designed a student network using depth-wise separable convolution. The teacher network learned the gunshot features with the guidance of the pre-trained large-scale teacher network. The proposed method was evaluated using the AudioSet dataset and the NIJ Grant 2016-DN-BX-0183 gunshot dataset and showed 95.6% and 83.5% accuracy, respectively.

The study [15] presented a novel approach to environmental sound classification using a newly collected dataset and a feature extraction function inspired by biological processes. The dataset comprises 5000 sounds which are classified into 50 classes. The proposed model involves feature generation utilizing a newly introduced lateral geniculate nucleus pattern (LGNPat), statistical moments, and discrete wavelet transform (DWT). The model uses loop-based neighborhood component analysis (INCA) for feature selection and k nearest neighbors (kNN) for classification with 10-fold cross-validation. The experimental results demonstrate a high classification accuracy of 93.34% when applying the proposed model to the ESC dataset.

In [16], a two-stage detection pipeline has been developed and tested for identifying gunshot sounds in tropical forests. The pipeline comprises an onboard detection algorithm and a spectrogram-based CNN. The classification pipeline aims to achieve high recall while tolerating increased false positives to assist human file annotation. To train the CNN, annotated data were used from two locations in Belize. The validation dataset consisting of 150 gunshots and 7044 background sounds yielded a recall of 0.95 and a precision of 0.85. The study indicates the possibility of using machine learning methods for detecting gunshots in tropical forests.

The authors introduced a hybrid ensemble classifier in [17] to classify environmental sounds by utilizing the UrbanSound8k dataset and MFCC features. An ensemble model [20] was developed by employing five distinct machine learning classifiers. The combination of all five classifiers produced the most desirable results, providing an accuracy of 79.4%. The approach proposed in this study has the potential to enhance the accuracy of ESC tasks. Similarly, [18] suggested a new technique for classifying environmental sounds, utilizing deep CNN. The method consists of several steps pre-processing, deep learning-driven feature extraction, feature concatenation, feature reduction, and classification. The short-time Fourier Transform (STFT) approach is utilized to transform the input sound signals into sound images, followed by pre-trained CNN models for feature extraction. The proposed method employs a support vector machine (SVM) classifier and its effectiveness is assessed on various datasets namely ESC-10, ESC-50, and UrbanSound8K, producing accuracy scores of 94.8%, 81.4%, and 78.14%, respectively.

The study [19] introduced a novel technique, called LMCC, to enhance environmental sound classification which fuses Log mel, log-scaled cochleagram, and log-scaled constant-Q transform features. The LMCC features are then fed into the CNN-GRUNN, a network comprising both a convolutional neural network and a gated recurrent unit neural network. The study conducted experiments using ESC-10, ESC-50, and UrbanSound8K datasets, and the results demonstrate that the proposed method exhibits high classification accuracy for all three datasets, namely ESC-10 (92.30%), ESC-50 (87.43%) and UrbanSound8K (96.10%).

Despite the contributions and superior results reported in the above-cited research papers, these studies lack in several

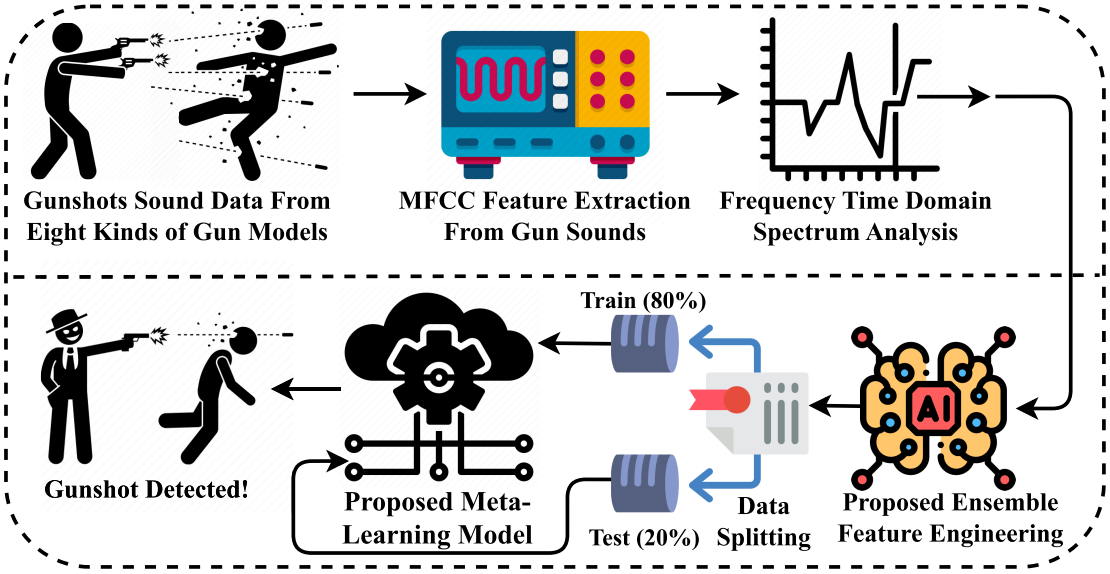

FIGURE 1. The gunshot detection methodology workflow analysis.

aspects. First, predominantly such studies focus on model building and improving performance by optimizing model architecture and parameters. So, the major emphasis is on the model engineering part, both for machine learning and deep learning models. Second, the studies that focus on the feature engineering part are very few, indicating that the feature engineering part is either ignored or under-studied. This study investigates feature engineering and proposes a novel feature extraction approach for a better approach. Additionally, the ensemble model is adopted for gunshot detection in view of the results reported for ensemble models. Thirdly, existing studies utilize classical machine learning and deep learning approaches. However, there is a growing need for more advanced techniques, such as meta-learning, to enhance the accurate recognition of gunshots and thereby contribute to crime prevention. Consequently, this study puts focus on meta-learning models and investigates their efficacy for gunshot detection.

III. PROPOSED METHODOLOGY

The gunshots sound dataset from eight kinds of gun models is utilized to conduct experiments in this study, as illustrated in Figure 1. The MFCC features are extracted from the signal data of gunshots. The frequency-time domain spectrum analysis is performed on gunshot sound data to understand the patterns of signals related to each target class. A novel ensemble feature engineering approach is proposed to create a new feature set extracted from MFCC features. The newly created features set is then divided into train and test portions with a ratio of 80:20. The 80% of data is utilized for training the proposed meta-learning MRK model, and 20% data is utilized to evaluate the performance. The novel proposed meta-learning-based MRK technique detects the gunshot sound with high-performance scores.

A. GUNSHOTS SOUNDS DATA

The gunshot sounds dataset from [21] is used for experiments. The dataset was originally collected from YouTube videos and contains 851 audio collected from public videos on YouTube from eight kinds of gun models. Each dataset audio segment lasts two seconds with a sampling frequency of 44.1 KHz. Each audio file was converted to wav file format. The dataset creators also validate the audio by carefully listening to them one by one. Figure 2 performs the histogram-based dataset distribution analysis. The analysis shows that the target class IMI Desert Eagle contains 100, M16 contains 100, M4 contains 100, MG-42 contains 100, MP5 contains 100, M249 contains 99, AK-12 contains 98, Zastava M92 contains 82, and AK-47 contains 72 samples. The analysis shows that data is imbalanced for only AK-47 and Zastava M92 classes.

B. MEL FREQUENCY CEPSTRAL COEFFICIENTS FEATURE EXTRACTION

MFCC features [22] are widely used in audio forensics for gunshot recognition. MFCC is a technique that represents the spectral envelope of an audio signal by extracting the Mel-scale frequency bands and calculating the cepstral coefficients of these bands. Gunshot signals have distinct characteristics that can be captured by the MFCC features, such as their short duration, high energy, and sharp attack. The first step is to extract the Mel-scaled filterbank features:

$$H_m[k] = \begin{cases} 0, & \text{for } k < f(m-1) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)}, & \text{for } f(m-1) \le k \le f(m) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)}, & \text{for } f(m) \le k < f(m+1) \\ 0, & \text{for } k > f(m+1) \end{cases} \quad (1)$$

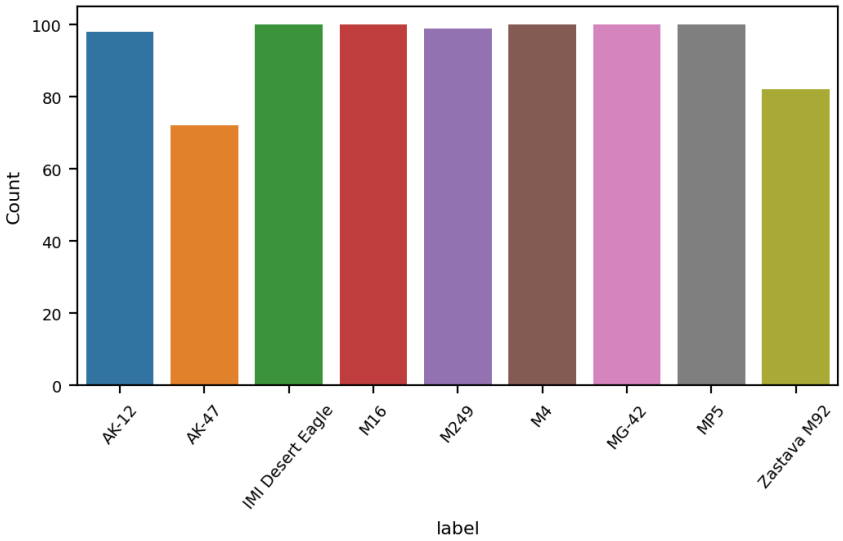

FIGURE 2. The histogram-based target Gunshots Sounds classes distributions analysis.

where k is the frequency index, m is the filter index, and $f(m)$ is the frequency in Hertz corresponding to the mth Mel frequency.

Next, we compute the logarithm of the energy in each filterbank using

$$E_m = \log\left(\sum_{k=0}^{N-1} |X(k)|^2 H_m[k]\right) \tag{2}$$

where $X(k)$ is the discrete Fourier transform (DFT) of the audio signal and N is the length of the DFT.

Afterward, we compute the MFCCs by applying a discrete cosine transform (DCT) to the filterbank energies using

$$c_n = \sum_{m=0}^{M-1} \cos\left[\frac{\pi}{M}(n+\frac{1}{2})m\right] E_m \tag{3}$$

where n is the MFCC index and M is the number of filterbanks.

Finally, we apply a lifter function to enhance the high-frequency components using

$$c_n = (1 + \frac{L}{2}\sin(\frac{\pi n}{L}))c_n \tag{4}$$

where L is the lifter parameter.

C. FREQUENCY TIME DOMAIN SPECTRUM ANALYSIS

Frequency time domain spectrum (FTDS) analysis using the extracted MFCC features from gunshots is illustrated in Figures 3 and 4. Each MFCC feature is plotted in time series corresponding to the target classes. The FTDS analysis demonstrates that all MFCC features have an amplitude in the range of 20 to 150 approximately. The analysis concludes that all MFCC features have a sharp peak at the frequency of the bullet crack typically characterizing the FTDS of a gunshot sound. The sudden release of energy causes this peak as the bullet breaks the sound barrier. The muzzle blast and the echo also have characteristic peaks in the FTDS.

D. NOVEL PROPOSED ENSEMBLE FEATURE ENGINEERING

A novel ensemble DWT-RFP feature engineering approach is proposed in this study. The MFCC features extracted from gunshot sound data are input to the DWT-RFP approach for feature extraction. The discrete wavelets transform [23] and random forest-based probabilistic [24] features are extracted from MFCC features. The newly created feature sets are combined and used to build applied machine learning and deep learning techniques.

1) PROBABILISTIC FEATURES

Probabilistic features are an important aspect of many machine learning models that help improve the accuracy of these predictions [25]. By incorporating probability distributions into their calculations, these models can account for uncertainty in the input data and produce more reliable predictions. Probabilistic features are an important tool for improving the accuracy and reliability of machine learning predictions. Let X be the input data set consisting of n samples, where each sample is a d-dimensional vector as follows

$$X = x_1, x_2, \ldots, x_n \tag{5}$$

Let Y be the corresponding target variable of the input data set, then

$$Y = y_1, y_2, \ldots, y_n \tag{6}$$

Let M be a classification model trained on the input data set X, which outputs the predicted class probabilities for each sample in X

$$M : X \rightarrow [0, 1]^k \tag{7}$$

where k is the number of classes.

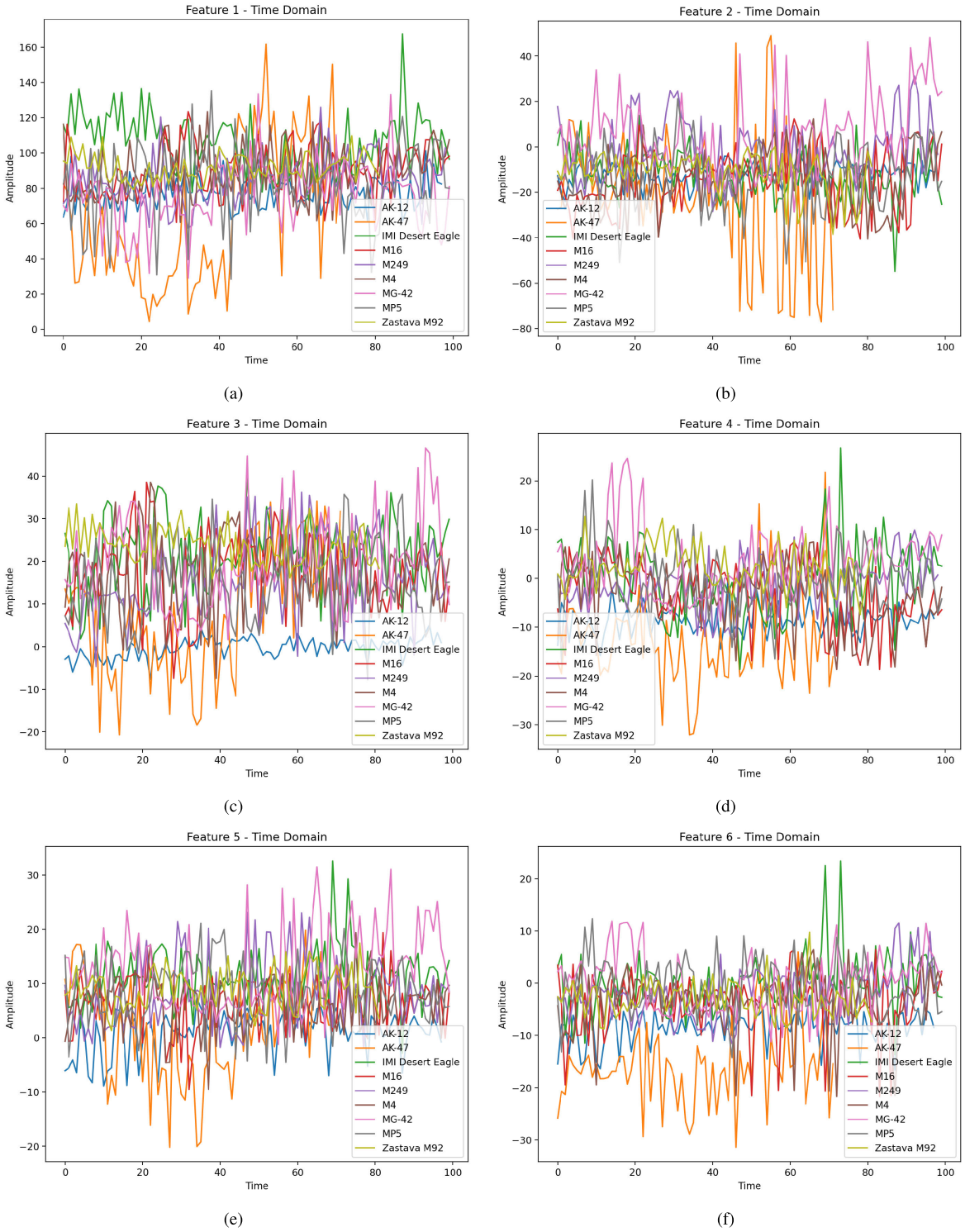

FIGURE 3. The frequency-time domain spectrum analysis of the first six dataset features regarding the target gunshot sounds classes.

For each sample x in X, the predicted class probabilities can be represented as follows

$$M(x) = p_1, p_2, \ldots, p_k \tag{8}$$

where p_i is the predicted probability of sample x belonging to class i.

The predicted probabilities can then be used as features to represent the input data set. Specifically, let P be a new data set obtained by extracting the predicted class probabilities from the original data set X

$$P = M(x_1), M(x_2), \ldots, M(x_n) \tag{9}$$

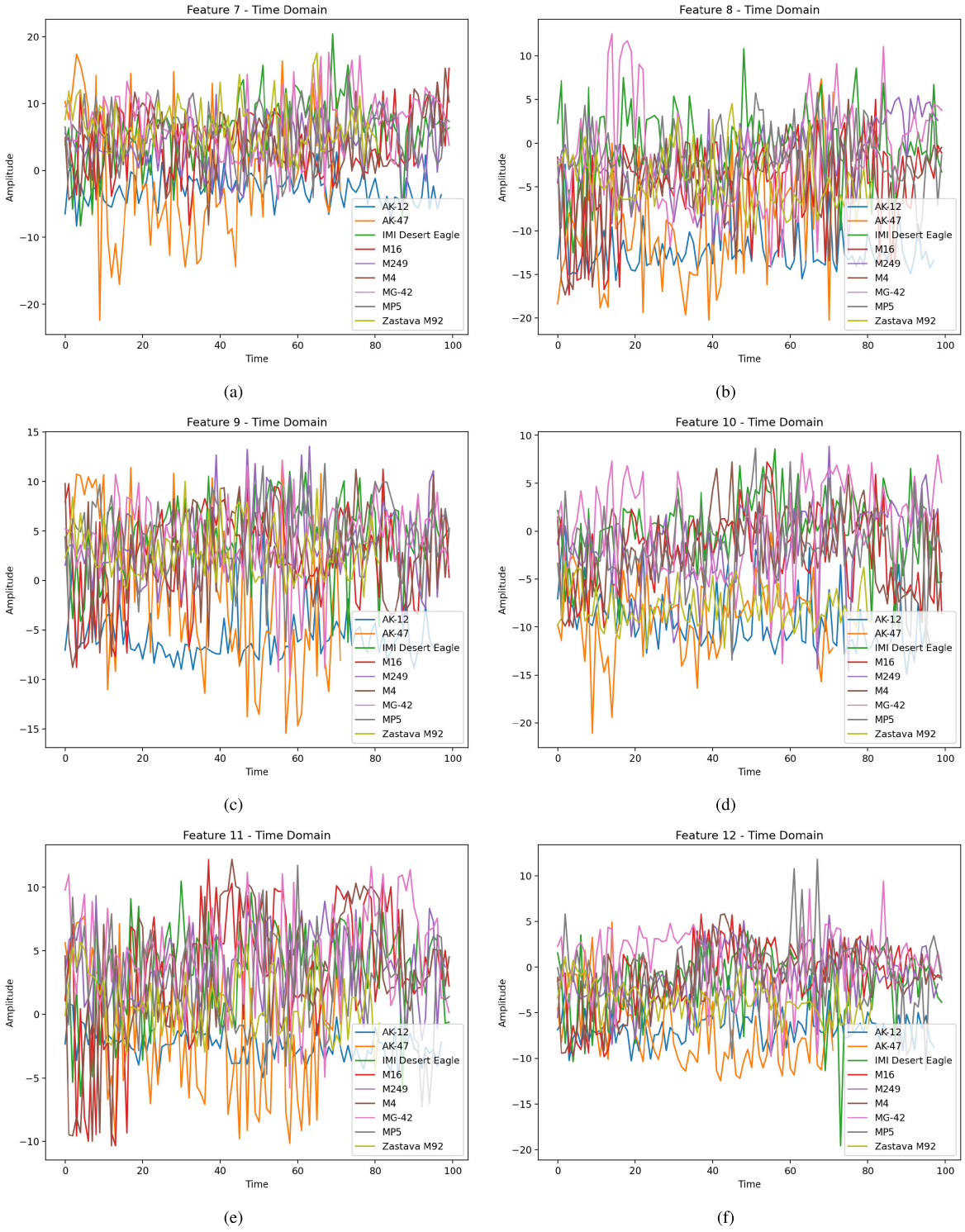

FIGURE 4. The frequency-time domain spectrum analysis of dataset features from seven indexes regarding the target gunshot sounds classes.

Each sample in P is a k-dimensional vector, where the i^{th} element of the vector is the predicted probability of the corresponding sample in X belonging to class i.

The resulting data set P can then be used as input for a subsequent machine learning model or analysis. Figure 5 shows the visual representation of the proposed ensemble feature engineering approach.

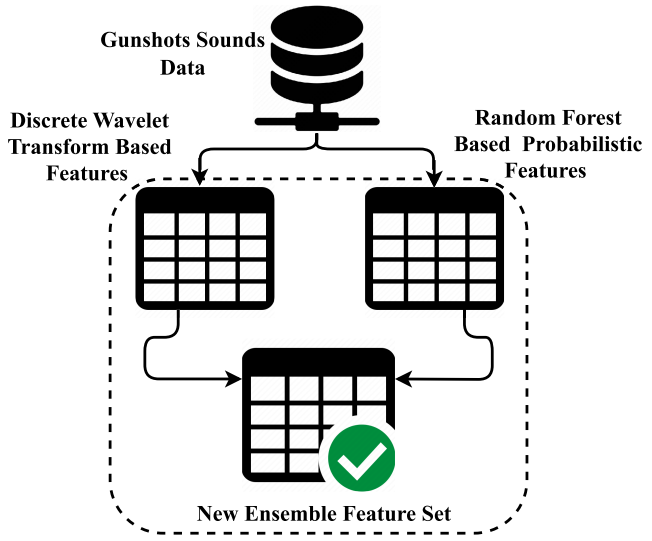

FIGURE 5. The workflow of the proposed ensemble feature engineering approach.

2) DISCRETE WAVELETS TRANSFORM

DWT is a widely used signal-processing technique for feature extraction in various fields. The DWT decomposes a signal into a set of wavelets representing different signal frequency components. Dilations and translations of a mother wavelet function generate the wavelets. The DWT can be computed using a filter bank approach, where the signal is passed through a set of high-pass and low-pass filters and then subsampled to obtain a down-sampled version of the signal. The DWT coefficients represent the wavelet coefficients at different scales and positions. The mathematical equations for the DWT are as follows.

Let $x[n]$ be a discrete-time signal of length N, and let $h[n]$ and $g[n]$ be the impulse response of the low-pass and high-pass filters, respectively. Then, the decomposition equation and reconstruction equations for DWT can be computed as

$$x[n] = \sum_{k=0}^{N-1} h[k] \cdot (2x[2n-k]) + \sum_{k=0}^{N-1} g[k] \cdot (2x[2n-k-1]) \tag{10}$$

$$x[n] = \sum_{k=0}^{N-1} \frac{1}{2} h[k] \cdot (x[2n-k] + x[2n-k-1]) + \sum_{k=0}^{N-1} \frac{1}{2} g[k] \cdot (x[2n-k] - x[2n-k-1]) \tag{11}$$

where n = 0, 1, …, N/2 - 1, and the DWT coefficients are defined as:

$$c_{j,k} = \frac{1}{2^j} \sum_{n=0}^{N-1} x[n] \cdot \psi^* j, k[n] \tag{12}$$

where j is the scale of the wavelet, k is the position of the wavelet, and $\psi^* j, k[n]$ is the complex conjugate of the wavelet function $\psi j, k[n]$.

The DWT coefficients can be used as features for various signal-processing tasks, such as signal denoising, signal compression, and pattern recognition. The DWT is widely used in image processing and computer vision applications, as well, due to its ability to extract both time-domain and frequency-domain information from signals.

Algorithm 1 shows the step-by-step flow of the proposed ensemble feature engineering approach.

Algorithm 1 DWT-RFP Algorithm

Input: MFCC features extracted from gunshot sound dataset.

Output: Newly created ensemble feature set for Gunshot Sound Detection | IMI Desert Eagle, M16, M4, MG-42, MP5, M249, AK-12, Zastava M92, or AK-47.

initiate;

1- $F_{dwt} \longleftarrow PYWT.WAVEDEC_{training}(TrS)$ // here TrS is the training set belonging to the original MFCC features data, and F_{dwt} are extracted DWT features.

2- $F_{rf} \longleftarrow RF_{training}(TrS)$ // here F_{rf} are the extracted probabilistic features.

3- $H_{features} \longleftarrow \sum\{F_{dwt} + F_{rf}\}$ // $H_{features} \epsilon$ New hybrid features set used for Gunshot Sound Detection.

end;

E. DATA SPLITTING

In this research, we utilized an 80:20 ratio for data split. The 80% portion of the data is used for training machine learning and deep learning methods while the remaining 20% is kept unseen to evaluate the performance of the employed models. The dataset splitting is carried out using the scikit-learn module's *train_test_split*() function. This approach aids in mitigating overfitting issues and enhancing the generalization capability of a model. Furthermore, we employed k-fold cross-validation data split to validate the obtained results.

F. APPLIED MACHINE AND DEEP LEARNING MODELS

In recent years, machine and deep learning techniques have been used for gunshot recognition in audio forensics [26]. The recognition of gunshots poses a significant challenge due to the variations in the acoustic characteristics of gunshots including the firearm's caliber, the distance between the shooter and the microphone, and the surrounding environment. The applications of these techniques in gunshot recognition have the potential to aid law enforcement agencies in investigating firearm-related crimes and provide valuable evidence in court proceedings. This section elaborates on the mathematical working of the applied machine and deep learning methods for gunshot recognition.

1) DECISION TREE CLASSIFIER

Decision tree classifier (DTC) is a popular approach to gunshot recognition [27]. It allows for the creation of a simple

and interpretable model that can classify audio events based on relevant features. These features may include spectral content, temporal characteristics, and other acoustic properties indicative of a gunshot. DTC can be trained on labeled datasets of gunshot and non-gunshot sounds, achieving high accuracy levels in identifying gunshots in new recordings. The mathematical notations to expressed DTC model are expressed as

$$\text{Information Gain} = \text{Entropy(parent)} - \sum_{i=1}^{k} \frac{n_i}{n} \text{Entropy(child}_i) \tag{13}$$

$$\text{Entropy} = -\sum_{i=1}^{c} p_i \log_2 p_i \tag{14}$$

$$\text{Gini Impurity} = 1 - \sum_{i=1}^{c} p_i^2 \tag{15}$$

where k is the number of children of the parent node, n_i is the number of instances in the i-th child, n is the total number of instances, c is the number of classes and p_i is the proportion of instances in the i-th class.

2) RANDOM FOREST CLASSIFIER

Random forest classifier (RFC) is a machine learning algorithm that has shown promising results for gunshot recognition in audio forensics [28]. RFC is an ensemble-based learning method that combines multiple decision trees to improve classification accuracy. The algorithm is trained on features extracted from the audio recordings, such as the gunshot signals' energy, duration, and spectral characteristics. The trained model can then be used to classify new audio recordings as either containing a gunshot or not. Let X be the input data and Y be the target variable. Given a set of training data

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n) \tag{16}$$

The algorithm builds a forest of decision trees by randomly selecting a subset of features for each tree and using a random sample of the training data to grow each tree. To predict the target variable Y for a new input data point x, the RFC combines the predictions of all the decision trees in the forest by taking the mode of the predicted classes. The RFC can be formulated mathematically as follows

- Let T be the number of decision trees in the forest.
- For each tree t in the forest, select a random subset of m features from the total set of p features.
- For each tree t, use a random sample of size n with replacement from the training data to grow the tree.
- For each tree t, at each node in the tree, select the best split among a random subset of k features from the m features selected in step 2.
- To predict the class of a new input data point x, let $f_t(x)$ be the predicted class for tree t. Then the predicted class for the RFC is

$$Y = mode(f_1(x), f_2(x), \ldots, f_T(x)) \tag{17}$$

3) K-NEIGHBORS CLASSIFIER

K-Neighbors classifier (KNC) is a widely used machine learning algorithm that can be trained to recognize and classify audio signals [29]. KNC operates on the principle of proximity-based classification, where it identifies the k-nearest neighbors of a given data point and assigns a label based on the majority class among those neighbors. For example,

- x: the input data point we want to classify
- X: the training data set
- y: the labels of the training data set
- K: the number of nearest neighbors to consider

The KNC can be described as follows

1. Calculate the Euclidean distance between x and each point in X:

$$d(x, X_i) = \sqrt{\sum_{j=1}^{p} (x_j - X_{ij})^2} \tag{18}$$

where p is the number of features in the data.

2. Select the K nearest neighbors based on the calculated distances.

3. Assign the class label of x as the majority class label among the K nearest neighbors:

$$y = \arg\max_{c \in C} \sum_{i=1}^{K} \mathbb{1}(y_i = c) \tag{19}$$

where C is the set of all possible class labels, and $\mathbb{1}(\cdot)$ is the indicator function that returns 1 if the condition is true and 0 otherwise.

4) LOGISTIC REGRESSION

Logistic regression (LR) is a popular statistical method that has been used in many different fields, including audio forensics [30]. LR models can be trained on a large dataset of known gunshot recordings to improve accuracy and reduce false positives. The use of logistic regression for gunshot recognition addresses potential issues related to sample bias and data collection challenges. The mathematical notations of the LR model are expressed as:

$$P(y = 1|\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}^\top \mathbf{x}}}, \tag{20}$$

where $\mathbf{x}$ is the input feature vector, $\mathbf{w}$ is the weight vector, and y is the binary output variable that takes on the value of 0 or 1.

The logistic function, also known as the sigmoid function, transforms the output of the linear function $\mathbf{w}^\top \mathbf{x}$ to a value between 0 and 1, which can be interpreted as the probability of the output variable y being 1 given the input feature vector $\mathbf{x}$. The cost function used in LR is the cross-entropy loss, which can be defined as

$$J(\mathbf{w}) = -\frac{1}{N} \sum_{i=1}^{N} y^{(i)} \log \left(P(y^{(i)} = 1|\mathbf{x}^{(i)}) \right) + (1 - y^{(i)}) \log \left(1 - P(y^{(i)} = 1|\mathbf{x}^{(i)}) \right), \tag{21}$$

where N is the number of training examples, $y^{(i)}$ is the ground truth label for the i-th example, and $\mathbf{x}^{(i)}$ is the input feature vector for the i-th example.

To train the LR, we can use gradient descent to minimize the cost function with respect to the weight vector $\mathbf{w}$. The update rule for gradient descent can be written as:

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \alpha \frac{\partial J(\mathbf{w}^{(t)})}{\partial \mathbf{w}}, \tag{22}$$

where α is the learning rate, and $\frac{\partial J(\mathbf{w}^{(t)})}{\partial \mathbf{w}}$ is the gradient of the cost function with respect to the weight vector $\mathbf{w}$, evaluated at the current weight vector $\mathbf{w}^{(t)}$.

5) GAUSSIAN NAIVE BAYES

Gaussian Naive Bayes (GNB) is a popular classification algorithm used in various fields including audio forensics [31]. GNB is based on the Bayes theorem and assumes that the features are independent. In audio forensics, the GNB classifier extracts audio features such as frequency, amplitude, and duration, which are then used to distinguish between gunshot and non-gunshot audio signals. GNB has proven to be a reliable and efficient algorithm for gunshot recognition in audio forensics. Let X be the input data set consisting of n samples, where each sample is a d-dimensional vector

$$X = x_1, x_2, \ldots, x_n \tag{23}$$

Let Y be the corresponding target variable of the input data set

$$Y = y_1, y_2, \ldots, y_n \tag{24}$$

The GNB model aims to predict the target variable Y based on the input data set X. The model assumes that each feature in X is conditionally independent given the class variable Y and that each feature follows a Gaussian distribution. The GNB model calculates the posterior probability of each class given the input features using Bayes' theorem

$$P(Y = c|X = x) = \frac{P(X = x|Y = c)P(Y = c)}{P(X = x)} \tag{25}$$

where c is a class label, and x is a sample in the input data set X.

The prior probability of each class $P(Y = c)$ can be estimated by the relative frequency of each class in the training data

$$P(Y = c) = \frac{\sum_{i=1}^{n}[y_i = c]}{n} \tag{26}$$

where$[y_i = c]$ is an indicator function that equals 1 if $y_i = c$, and 0 otherwise.

The likelihood of the input features P(X = x | Y = c) can be modeled using a Gaussian distribution with mean $mu_candvariancesigma_c^2$:

$$P(X = x|Y = c) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi\sigma_{c,i}^2}} e^{-\frac{(x_i-\mu_{c,i})^2}{2\sigma_{c,i}^2}} \tag{27}$$

where $mu_candsigma_c^2$ are the mean and variance of the i-th feature in class c, respectively.

The model selects the class with the highest posterior probability as the predicted class for each sample in the input data set X

$$\hat{y} = \underset{c}{\text{argmax}}\ P(Y = c|X = x) \tag{28}$$

where $\hat{y}$ is the predicted class label.

The GNB model is a simple yet effective classification model, which is especially useful for high-dimensional data sets with continuous features.

6) SUPPORT VECTOR MACHINE

SVM is a machine learning algorithm widely used in various applications including audio classification [32], natural language processing [33], etc. Researchers have explored the use of linear SVM for Gunshot audio classification which has significant implications for public safety and security. Linear SVM works by finding a hyperplane that separates the data into two classes, with the maximum margin between the two classes. The decision function of a linear SVM classifier can be represented as:

$$f(x) = w \cdot x + b \tag{29}$$

where x is the input vector, $\mathbf{w}$ is the weight vector, b is the bias, and $\cdot$ denotes the dot product.

The goal of the SVM algorithm is to find the optimal weight vector $\mathbf{w}$ and bias b that maximizes the margin, which is the distance between the decision boundary and the closest data points of each class. The optimization problem can be formulated as:

$$\underset{w,b}{\text{minimize}} \frac{1}{2}|w|^2 \tag{30}$$

subject to $y_i(w \cdot x_i + b) \geq 1$ for all $i = 1, 2, \ldots, n$

where $|w|^2$ is the squared norm of the weight vector, and y_i is the target label of input vector x_i.

This is a quadratic programming problem that can be solved using standard optimization techniques. Once the optimal weight vector $\mathbf{w}$ and bias b are obtained, new data points can be classified as follows:

$$\hat{y} = \text{sign}(f(x)) = \text{sign}(w \cdot x + b) \tag{31}$$

where $\hat{y}$ is the predicted label and sign is the sign function that returns $+1$ for positive values and -1 for negative values.

Overall, the linear SVM aims to find the optimal decision boundary that separates the input data points into two different classes by maximizing the margin between them.

7) ADABOOST CLASSIFIER

AdaBoost Classifier (ABC) effectively detects and classifies various gunshots including indoor and outdoor gunshots, suppressed and unsuppressed gunshots, and single and multiple gunshots [34]. ABC trains a series of weak classifiers and combines their predictions to form a strong classifier. Misclassified samples are assigned higher weights to prioritize

the classification of more challenging samples. The final model is a weighted sum of the weak classifiers, with the weight of each classifier based on its performance in the training process. First, the ABC assigns equal weights to each data point in the training set

$$w_i = \frac{1}{N} \quad \text{for } i = 1, 2, \ldots, N \tag{32}$$

where w_i is the weight assigned to the ith data point and N is the number of data points in the training set.

Then, it fits a weak learner, such as a decision tree, to the training data, giving more weight to the misclassified data points:

$$h_t(x) = \begin{cases} 1, & \text{if } \sum_{i=1}^{N} w_i y_i h_t(x_i) \geq \frac{1}{2} \sum_{i=1}^{N} w_i \ - 1, \\ 0, & \text{otherwise} \end{cases} \tag{33}$$

where $h_t(x)$ is the output of the weak learner at iteration t, x_i is the ith data point in the training set, y_i is the corresponding label, and t is the current iteration.

Next, the classifier calculates the weighted error rate of the weak learner

$$\epsilon_t = \sum_{i=1}^{N} w_i I(y_i \neq h_t(x_i)) \tag{34}$$

where I is the indicator function.

Based on the weighted error rate, the ABC calculates the weight of the weak learner in the final ensemble

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right) \tag{35}$$

Finally, it updates the weights of the data points in the training set, giving more weight to the misclassified data points and less weight to the correctly classified data points

$$w_i \leftarrow w_i \exp(-\alpha_t y_i h_t(x_i)) \tag{36}$$

The process is repeated for a specified number of iterations or until the desired accuracy is achieved. The final output of the classifier is a weighted sum of the weak learners

$$H(x) = \text{sign} \left(\sum_{t=1}^{T} \alpha_t h_t(x) \right) \tag{37}$$

where T is the total number of weak learners in the ensemble and $H(x)$ is the predicted label for input x.

8) RECURRENT NEURAL NETWORK

Recurrent Neural Networks (RNNs) have been extensively studied for sequential data analysis [35], and their potential for gunshot audio classification has garnered significant interest. An RNN for gunshot audio classification involves processing sequential audio data by using the network's hidden state to capture temporal dependencies in the audio data. The RNN comprises input, hidden, and output layers. Each layer consists of neurons that are connected to neurons in previous and subsequent layers through weighted connections. The working flow of an RNN model can be represented mathematically as follows.

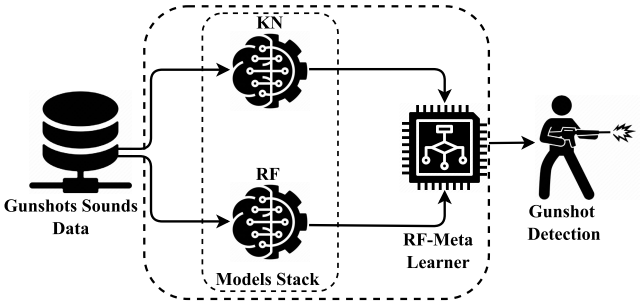

FIGURE 6. The novel proposed meta-learning-based stacked model analysis.

Let x_t be the input at time step t, h_t be the hidden state at time step t, and y_t be the output at time step t. The hidden state h_t is computed based on the input x_t and the previous hidden state h_{t-1} as follows

$$h_t = f(W_{xh} x_t + W_{hh} h_{t-1} + b_h) \tag{38}$$

where W_{xh} is the weight matrix connecting the input x_t to the hidden state h_t, W_{hh} is the weight matrix connecting the previous hidden state h_{t-1} to the current hidden state h_t, and b_h is the bias term.

The output y_t is computed based on the current hidden state h_t as follows

$$y_t = g(W_{hy} h_t + b_y) \tag{39}$$

where W_{hy} is the weight matrix connecting the hidden state h_t to the output y_t, and b_y is the bias term.

During training, the parameters of the RNN model including the weight matrices W_{xh}, W_{hh}, and W_{hy}, as well as the bias, terms b_h and b_y, are updated using backpropagation through time.

G. PROPOSED META-LEARNING BASED STACKED MODEL

A novel meta-learning-based Meta-RF-KN (MRK) is proposed to detect the gunshots from newly created ensemble features extracted from MFCC features of gunshot sound data, as shown in Figure 6. The newly created ensemble features data is input to the KN and RF techniques for predictions. The output predictions from both stacked classifiers are combined and input to a random forest-based meta-learner for the final prediction. The final forest-based meta-learner uses the strength of each model in the stack by using the output as input. The experimental results show that the proposed MRK approach achieves high performance for gunshot detection.

H. HYPERPARAMETER SETTINGS

Table 2 hyperparameter optimization analysis based on best fit selected features for applied machine learning and deep learning techniques. The recursive mechanisms of training and validation determine the best-fit hyperparameters in this

TABLE 2. Hyperparameter optimization for machine learning and deep learning models.

Model	Parameters and description
DTC	max_depth=300, criterion='gini', splitter='best'
RFC	n_estimators=300,max_depth=300, random_state=0, criterion='gini'
KNC	n_neighbors=2, weights='uniform', leaf_size=30
LR	random_state=0,max_iter=500,solver='liblinear'
GNB	var_smoothing=1e-9
SVC	random_state=0,max_iter=50, penalty='l2'
ABC	n_estimators=100, algorithm='SAMME.R', learning_rate=1.0
RNN	loss = 'categorical_crossentropy',optimizer = 'adam' ,metrics='accuracy', activation='softmax'

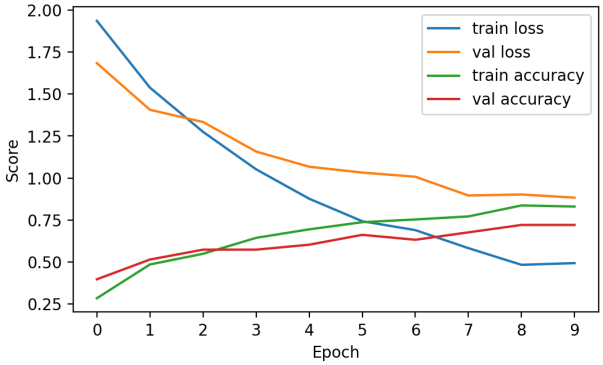

FIGURE 7. Loss and accuracy of applied deep learning based RNN model during training with MFCC features.

analysis. The hyperparameter tuning validated each applied technique's performance and achieved a high accuracy score for gunshot detection.

IV. RESULTS AND DISCUSSIONS

This research aims to explore the results and discussions of machine learning models employed for gunshot sound detection. The proposed study utilized a diverse dataset comprising gunshot sounds recorded under various conditions to evaluate the performance of applied models. The results and discussions of this research highlighted the effectiveness of machine learning models for gunshot sound classification.

A. EXPERIMENTAL SETUP

The experimental setup for conducting research experiments is analyzed in this section. The dataset is split into 80% training and 20% testing. The Google Colab [36] environment is used to conduct the experiments. The used environment is based on a GPU backend with 13 GB RAM and 90 GB of disk space. The performance evaluation metrics for gunshot recognition are accuracy score, precision score, recall score, F1 score, time series analysis, confusion matrix, standard deviation, and computational complexity.

B. RESULTS WITH MFCC FEATURES

Training loss and accuracy and validation loss and accuracy of deep learning models are illustrated in Figure 7. During the training of the RNN method with MFCC features, each epoch evaluates the loss and accuracy scores. The analysis demonstrates a high train and validation loss score from epochs 1 to 4. After that, the RNN model adjusts its optimal weights, which reduces the loss score and improves the performance accuracy scores till the last epoch of training.

TABLE 3. The testing performance analysis of applied methods with MFCC features.

Technique	Accuracy	Precision	Recall	F1
DTC	0.53	0.57	0.58	0.57
RFC	0.65	0.68	0.70	0.69
KNC	0.58	0.64	0.63	0.60
LR	0.54	0.57	0.59	0.56
GNB	0.50	0.49	0.55	0.51
SVC	0.42	0.43	0.46	0.40
ABC	0.18	0.12	0.17	0.10
RNN	0.63	0.65	0.68	0.66
MRK	**0.85**	**0.87**	**0.87**	**0.87**

TABLE 4. The class-wise performance analysis of proposed MRK approach with MFCC features.

Target Class	Precision	Recall	F1
AK-12	0.89	1.00	0.94
AK-47	1.00	1.00	1.00
IMI Desert Eagle	0.75	0.90	0.82
M16	0.91	0.84	0.87
M249	0.88	0.74	0.80
M4	0.94	0.73	0.82
MG-42	0.95	0.86	0.90
MP5	0.62	0.80	0.70
Zastava M92	0.93	0.93	0.93
Average	**0.87**	**0.87**	**0.87**

The performance analysis of applied machine and deep learning techniques with MFCC features is given in Table 3. The applied ABC and SVC technique achieved poor accuracy scores of 0.18 and 0.42, respectively. The analysis shows that using MFCC features, low-performance scores are achieved by machine learning and deep learning models. The proposed meta-learning model MRK achieved an acceptable score. However, there is a need for further improvement to achieve high-performance scores for gunshot detection.

Table 4 contains the target class-wise performance analysis of the proposed meta-learning-based MRK approach with MFCC features. The analysis shows that the MRK approach achieves 100% precision, recall, and f1 scores for AK-47 gunshot detection. Only the low-performance scores are for MP5 gunshot detection. This analysis indicates that the proposed meta-learning model achieved good performance scores with MFCC features for gunshot detection.

The confusion matrix for all applied techniques using MFCC features is illustrated in Figure 8. Each applied method prediction label and the true label-based confusion matrix are analyzed in this regard. The analysis demonstrates that using the MFCC features, the applied machine learning and deep learning techniques achieved a high error rate during the gunshot classification.

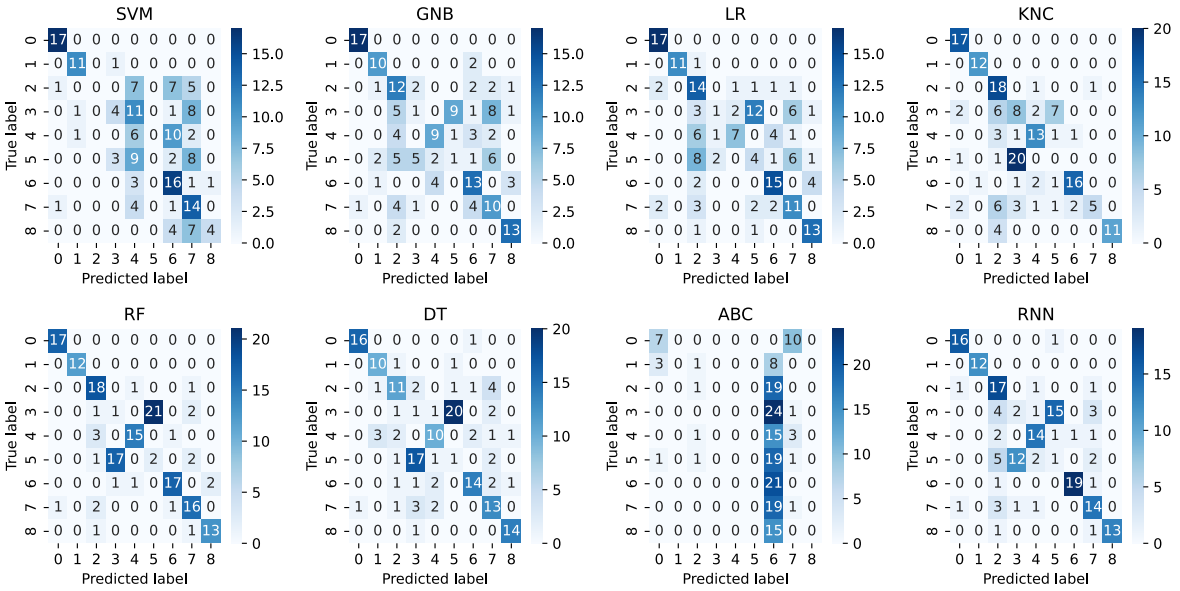

FIGURE 8. The confusion matrix analysis of all applied techniques using MFCC features.

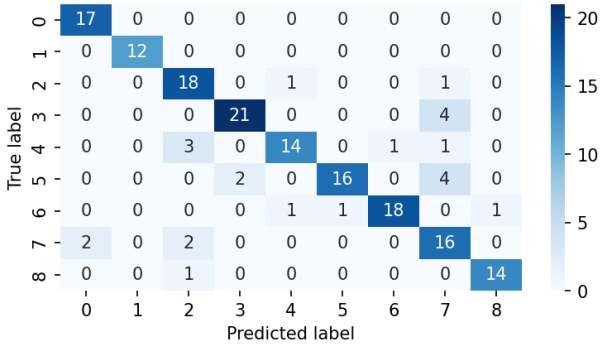

FIGURE 9. The confusion matrix of proposed MRK approach with MFCC features.

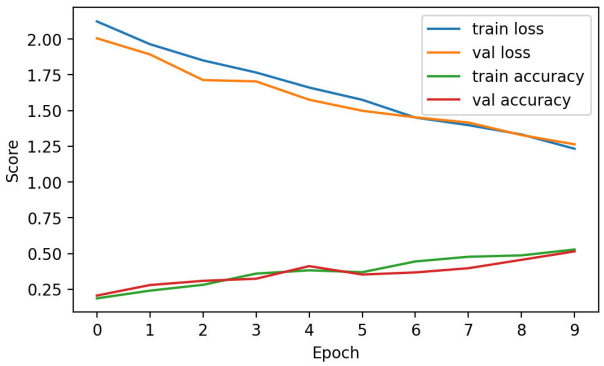

FIGURE 10. Training and testing performance of RNN model during training with ensemble features.

The confusion matrix of the proposed MRK approach with MFCC features is shown in Figure 9. The analysis shows that the proposed approach has higher performance regarding the number of correct predictions compared to other applied models. The analysis indicates that the proposed meta-learning model achieves a low error rate for gunshot detection.

C. RESULTS WITH PROPOSED ENSEMBLE FEATURES

Figure 10 shows the training and testing accuracy and loss for applied deep learning models using the ensemble features. During the training of the RNN method with ensemble features, each epoch evaluates the loss and accuracy scores. The analysis demonstrates a high train and validation loss score from epochs 1 to 5. After that, the RNN model adjusts its optimal weights, which reduces the loss score and improves the performance accuracy scores till the last epoch of training. The RNN model achieved poor performance scores for gunshot detection in this analysis.

The performance analysis of applied machine and deep learning techniques with proposed ensemble features is carried out and results are given in Table 5. The applied SVC and RNN techniques achieved poor accuracy scores of 0.46 and 0.45, respectively while DTC and RFC achieves better performance with a 0.76 accuracy score each. Results indicate that the performance of the models is elevated when using the proposed ensemble features compared to previous results with MFCC features alone. The best performance is achieved by the proposed meta-learning model MRK with an accuracy score of 0.96. Overall, the performance of applied methods is improved for gunshot detection using the proposed ensemble features.

Table 6 contains the target class-wise performance analysis of the proposed meta-learning-based MRK approach with ensemble features. The analysis shows that the MRK approach achieves 100% precision, recall, and F1 scores for AK-12, AK-47, IMI Desert Eagle, M249, MG-42, MP5, and

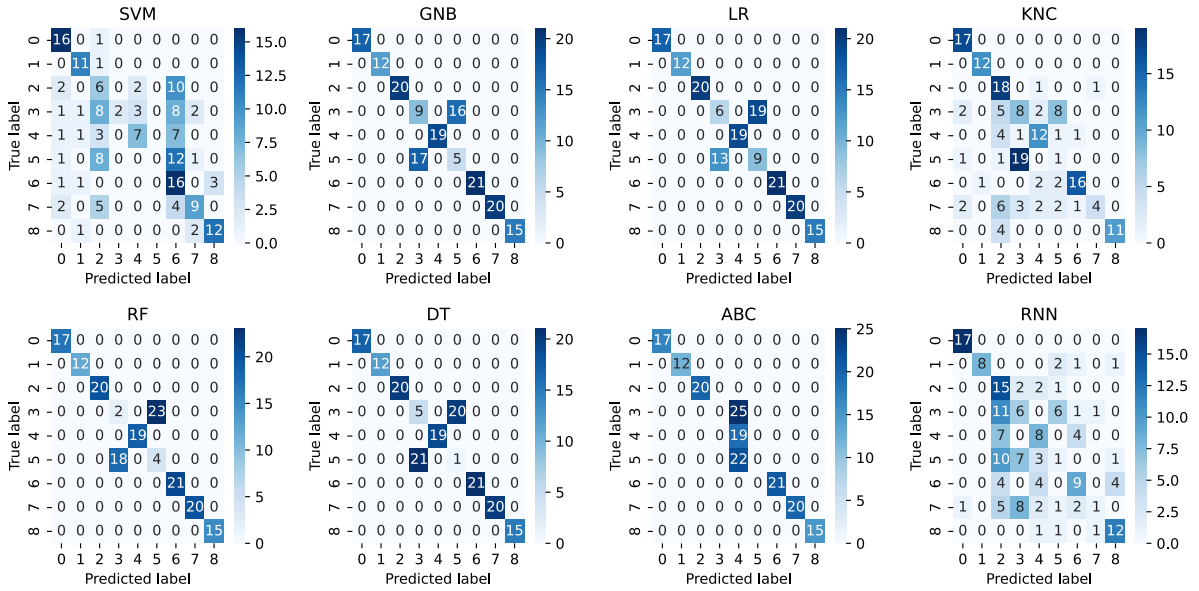

FIGURE 11. The confusion matrices of all applied techniques using ensemble features.

TABLE 5. Testing performance of employed models with the proposed feature engineering approach.

Technique	Accuracy	Precision	Recall	F1	AUC
DTC	0.76	0.80	0.81	0.80	0.89
RFC	0.76	0.81	0.81	0.81	0.96
KNC	0.58	0.65	0.62	0.60	0.82
LR	0.81	0.85	0.85	0.85	0.97
GNB	0.81	0.84	0.84	0.84	0.97
SVC	0.46	0.54	0.51	0.46	0.97
ABC	0.73	0.65	0.73	0.66	0.98
RNN	0.45	0.50	0.49	0.47	0.84
MRK	**0.96**	**0.97**	**0.97**	**0.97**	**0.99**

TABLE 6. The class-wise performance analysis of proposed MRK approach with proposed ensemble features.

Target Class	Precision	Recall	F1
AK-12	1.00	1.00	1.00
AK-47	1.00	1.00	1.00
IMI Desert Eagle	1.00	1.00	1.00
M16	0.88	0.88	0.88
M249	1.00	1.00	1.00
M4	0.86	0.86	0.86
MG-42	1.00	1.00	1.00
MP5	1.00	1.00	1.00
Zastava M92	1.00	1.00	1.00
Average	**0.97**	**0.97**	**0.97**

Zastava M92 gunshot detection. However, several classes have lower precision, recall, and F1 scores. For example, M16 and M4 classes have the precision of 0.88 and 0.86, respectively and their recall and F1 scores are also the lowest among other classes. Despite that average precision of 0.97 can be obtained for 9 classes using the proposed ensemble features with the proposed MRK model.

The confusion matrix analysis of all applied models using ensemble features is illustrated in Figure 11. Each applied

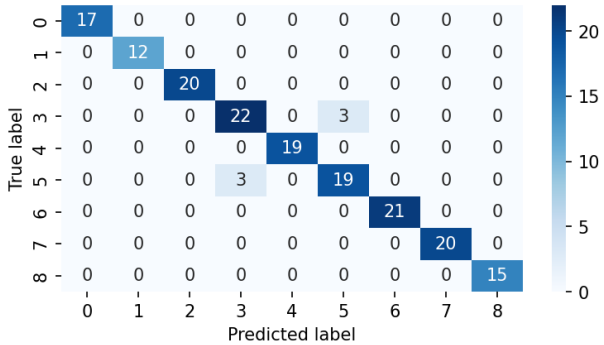

FIGURE 12. The confusion matrix of proposed MRK approach with ensemble features.

method prediction label and the true label-based confusion matrix are analyzed. The analysis demonstrates that machine learning and deep learning techniques achieved a minimum number of wrong predictions for gunshot classification when the proposed ensemble features are used for training the models. The performance of the models is significantly enhanced when ensemble features are used. Results show the efficacy of the proposed ensemble features for gunshot classification.

The confusion matrix analysis of the proposed MRK approach with ensemble features is visualized in Figure 12. The analysis indicates that the proposed MRK model obtains the highest number of correct predictions, with only 6 wrong predictions, when used with the proposed ensemble features. These predictions are much better than those of MFCC-based prediction with the MRK model. We can say that using the ensemble features, the proposed meta-learning model achieves a minimum error rate for gunshot detection.

TABLE 7. Performance analysis of proposed MRK model using k-fold cross-validation.

Techniques	K-folds	With MFCC Features		With New Ensemble Features	
		Kfold- Accuracy	Standard Deviation (+/-)	Kfold- Accuracy	Standard Deviation (+/-)
DTC	10	0.55	0.0402	0.77	0.0506
RFC	10	0.63	0.0561	0.77	0.0372
KNC	10	0.52	0.0400	0.52	0.0487
LR	10	0.56	0.0635	0.81	0.0312
GNB	10	0.52	0.0339	0.81	0.0314
SVC	10	0.44	0.0498	0.43	0.0726
ABC	10	0.17	0.0462	0.72	0.1110
RNN	10	0.21	0.2591	0.08	0.0947
MRK	**10**	**0.86**	**0.0287**	**0.99**	**0.0092**

D. K-FOLD CROSS-VALIDATION

Results for performance validation of applied machine learning and deep learning techniques based on K-fold cross-validation are given in Table 7. The k-fold Validation helps obtain a more robust estimate of the model's performance by reducing the impact of data variability and provides a better assessment of the model's generalization ability. The cross-validation analysis shows that using the MFCC features low-performance scores are achieved with high standard deviation. The proposed meta-learning model achieved acceptable k-fold scores with MFCC features. The analysis demonstrates that by using the proposed ensemble features the performance of the applied technique is improved. The analysis concludes that using the ensemble features proposed MRK approach achieved a 0.99 performance score with a minimal standard deviation score of 0.0092.

E. RESULTS FOR COMPUTATIONAL COMPLEXITY

All employed models are evaluated regarding the time complexity and results are given in Table 8. Results indicate that high computational time is needed when MFCC features are used for training the models. On the other hand, using the proposed ensemble features, the applied techniques require less training time and thus have lower computational complexity. Similarly, the deep learning-based RNN model has high computational complexity with MFCC features while requiring less training time when the ensemble features are used. The proposed MRK technique has the highest training time of 12.01 sec with the ensemble features, compared to 6.98 sec with MFCC features. However, it obtains the highest accuracy score of 0.99 using ensemble features.

F. COMPARISON WITH STATE-OF-THE-ART APPROACHES

The comparison with state-of-the-art approaches is performed and results are given in Table 9. We have compared the proposed approach with previously published studies from the year 2021 to 2022. The previous authors mainly used classical machine learning and deep learning techniques for sound classification. The analysis shows that the proposed approach outperformed existing models with a high accuracy score of 0.99 for detecting gunshot sounds. We have covered the research gap related to low-performance scores in previously published studies.

TABLE 8. The computational complexity of applied models.

Technique	Time (Seconds)	
	With MFCC features	With ensemble features
DTC	0.02	0.04
RFC	1.80	1.47
KNC	0.02	0.006
LR	0.30	0.16
GNB	0.01	0.01
SVC	0.07	0.13
ABC	0.59	1.14
RNN	11.95	8.42
MRK	6.98	12.01

TABLE 9. The performance comparisons with state-of-the-art approaches for gunshot sound classification.

Ref.	Year	Model	Accuracy
[1]	2022	Convolutional Neural Network	0.90
[4]	2022	Transformers Model	0.93
[21]	2021	K-Nearest Neighbors	0.94
[37]	2022	Dense Neural Network	0.95
Current study	**2023**	**MRK**	**0.99**

G. ADVANTAGES AND LIMITATIONS OF PROPOSED APPROACH

The proposed research is focused on the recognition of gunshot sounds which serve as crucial evidence in criminal investigations. The accurate gunshot detection would enable law enforcement officials to ascertain various details such as the count of fired shots, the caliber of the firearm used for firing, and the proximity from which the shots originated. The proposed research holds the potential to enhance gunshot detection accuracy, assisting to take timely measures and contributing to preventing crimes. The focus of this study is to enhance the accuracy and efficiency of gunshot detection by employing a meta-learning model with a novel feature set. Despite the fact that the study provides highly accurate results and superior performance compared to existing state-of-the-art gunshot detection models, the proposed approach has several limitations.

- The computational complexity of the proposed approach is high due to two factors. First, the proposed feature engineering approach utilizes probabilistic output from multiple approaches which increases its computational time. And secondly, the gunshot detection model is a

meta-learning model involving base and meta-models. Obtaining output from multiple models and providing the final prediction also leads to increased computational complexity. We intend to reduce the time complexity of the proposed technique in future work.

- The background noise in real-world environments, such as sirens, traffic, or conversations, can interfere with the accurate recognition of gunshot sounds. Similarly, changes in the indoor settings, place of the gunshot fires, the complexity of outdoor noise factors, and simultaneous firing from multiple guns are not considered. For an efficient approach, these factors must be incorporated in the approach. We intend to extend this research further to cover these aspects in the future.

V. CONCLUSION AND FUTURE WORK

Gunshot sound classification is an important element of crime investigation and timely detection can be used to stop and handle crimes. Gunshot sounds detection using a meta-learning approach is investigated in this study. A benchmark gunshot sounds dataset containing 851 audio samples collected from public videos on YouTube from eight kinds of gun models, is used to conduct experiments. A novel DWT-RFP feature engineering approach is proposed, which takes MFCC features extracted from gunshot sound data as input for feature extraction. A novel MRK model is proposed to detect gunshots based on newly created ensemble features with a DWT-RFP approach. Seven advanced machine learning techniques and one deep learning model are applied for gunshot detection. Extensive experiments show that the proposed MRK approach achieved a 0.99 k-fold accuracy score and outperforms existing state-of-the-art models for gunshot sound classification. In the proposed research study, we have achieved high-performance sores compared to state-of-the-art approaches. However, the computation complexity of the proposed technique is high. We intend to reduce the time complexity of the proposed technique in future work. We also want to enhance the dataset by collecting audio of more guns and analyzing the performance of the proposed model with a higher number of classes.

REFERENCES

[1] S. Raponi, G. Oligeri, and I. M. Ali, "Sound of guns: Digital forensics of gun audio samples meets artificial intelligence," *Multimedia Tools Appl.*, vol. 81, no. 21, pp. 30387–30412, Sep. 2022.

[2] B. Tardif, D. Lo, and R. Goubran, "Gunshot sound measurement and analysis," in *Proc. IEEE Sensors Appl. Symp. (SAS)*, Aug. 2021, pp. 1–6.

[3] D. Mazeika, "The effect of unreported gun-related violent crime on crime hot spots," *Secur. J.*, vol. 36, no. 1, pp. 1–17, Feb. 2022.

[4] R. Nijhawan, S. A. Ansari, S. Kumar, F. Alassery, and S. M. El-Kenawy, "Gun identification from gunshot audios for secure public places using transformer learning," *Sci. Rep.*, vol. 12, no. 1, p. 13300, Aug. 2022.

[5] J. R. Silva, "Global mass shootings: Comparing the United States against developed and developing countries," *Int. J. Comparative Appl. Criminal Justice*, vol. 2022, pp. 1–24, Mar. 2022.

[6] D. Watson, L. Howes, S. Dinnen, M. Bull, and S. N. Amin, "Trends in and social dynamics of crime in the Pacific," in *Policing in the Pacific Islands*. Cham, Switzerland: Springer, 2023, pp. 37–82.

[7] A. Raza, K. Munir, and M. Almutairi, "A novel deep learning approach for deepfake image detection," *Appl. Sci.*, vol. 12, no. 19, p. 9820, Sep. 2022.

[8] Z. Chen, H. Zheng, J. Huang, L. Wu, S. Cheng, Q. Zhou, and Y. Yang, "A wireless gunshot recognition system based on tri-axis accelerometer and lightweight deep learning," *IEEE Internet Things J.*, early access, May 8, 2023, doi: 10.1109/JIOT.2023.3273859.

[9] L. Xie, C. Lu, Z. Liu, W. Chen, Y. Zhu, and T. Xu, "The evaluation of automobile interior acceleration sound fused with physiological signal using a hybrid deep neural network," *Mech. Syst. Signal Process.*, vol. 184, Feb. 2023, Art. no. 109675.

[10] S. Agarwal, K. Khatter, and D. Relan, "Security threat sounds classification using neural network," in *Proc. 8th Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, Mar. 2021, pp. 690–694.

[11] A. Greco, N. Petkov, A. Saggese, and M. Vento, "AReN: A deep learning approach for sound event recognition using a brain inspired representation," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3610–3624, 2020.

[12] M. Massoudi, S. Verma, and R. Jain, "Urban sound classification using CNN," in *Proc. 6th Int. Conf. Inventive Comput. Technol. (ICICT)*, Jan. 2021, pp. 583–589.

[13] M. G. Ragab, S. J. Abdulkadir, N. Aziz, H. Alhussian, A. Bala, and A. Alqushaibi, "An ensemble one dimensional convolutional neural network with Bayesian optimization for environmental sound classification," *Appl. Sci.*, vol. 11, no. 10, p. 4660, May 2021.

[14] J. Li, J. Guo, X. Sun, C. Li, and L. Meng, "A fast identification method of gunshot types based on knowledge distillation," *Appl. Sci.*, vol. 12, no. 11, p. 5526, May 2022.

[15] B. Taşcı, M. R. Acharya, P. D. Barua, A. M. Yildiz, M. V. Gun, T. Keles, S. Dogan, and T. Tuncer, "A new lateral geniculate nucleus pattern-based environmental sound classification using a new large sound dataset," *Appl. Acoust.*, vol. 196, Jul. 2022, Art. no. 108897.

[16] L. K. D. Katsis, A. P. Hill, E. Piña-Covarrubias, P. Prince, A. Rogers, C. P. Doncaster, and J. L. Snaddon, "Automated detection of gunshots in tropical forests using convolutional neural networks," *Ecolog. Indicators*, vol. 141, Aug. 2022, Art. no. 109128.

[17] A. Bansal and N. K. Garg, "Environmental sound classification using hybrid ensemble model," *Proc. Comput. Sci.*, vol. 218, pp. 418–428, Jan. 2023.

[18] F. Demir, M. Turkoglu, M. Aslan, and A. Sengur, "A new pyramidal concatenated CNN approach for environmental sound classification," *Appl. Acoust.*, vol. 170, Dec. 2020, Art. no. 107520.

[19] Y. Zhang, J. Zeng, Y. Li, and D. Chen, "Convolutional neural network-gated recurrent unit neural network with feature fusion for environmental sound classification," *Autom. Control Comput. Sci.*, vol. 55, no. 4, pp. 311–318, Jul. 2021.

[20] A. Raza, F. Rustam, H. U. R. Siddiqui, I. D. L. T. Diez, and I. Ashraf, "Predicting microbe organisms using data of living micro forms of life and hybrid microbes classifier," *PLoS ONE*, vol. 18, no. 4, Apr. 2023, Art. no. e0284522.

[21] T. Tuncer, S. Dogan, E. Akbal, and E. Aydemir, "An automated gunshot audio classification method based on finger pattern feature generator and iterative relieff feature selector," *Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi*, vol. 8, no. 14, pp. 225–243, 2021.

[22] Q. Yao, Y. Wang, Y. Yang, and Y. Shi, "Seal call recognition based on general regression neural network using Mel-frequency cepstrum coefficient features," *EURASIP J. Adv. Signal Process.*, vol. 2023, no. 1, p. 48, May 2023.

[23] O. Das and D. Bagci Das, "Smart machine fault diagnostics based on fault specified discrete wavelet transform," *J. Brazilian Soc. Mech. Sci. Eng.*, vol. 45, no. 1, p. 55, Jan. 2023.

[24] A. Raza, H. U. R. Siddiqui, K. Munir, M. Almutairi, F. Rustam, and I. Ashraf, "Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction," *PLoS ONE*, vol. 17, no. 11, Nov. 2022, Art. no. e0276525.

[25] A. Raza, F. Rustam, H. U. R. Siddiqui, I. D. L. T. Diez, B. Garcia-Zapirain, E. Lee, and I. Ashraf, "Predicting genetic disorder and types of disorder using chain classifier approach," *Genes*, vol. 14, no. 1, p. 71, Dec. 2022.

[26] V. Singh, K. C. Ray, and S. Tripathy, "Robust gunshot features and its classification using support vector machine for wildlife protection," in *Proc. SIC*. Cham, Switzerland: Springer, 2020, pp. 939–948.

[27] V. Rinsha and G. Jagadanand, "Rolling average–decision tree–based fault detection of neutral point clamped inverters," *IEEE J. Emerg. Sel. Topics Ind. Electron.*, vol. 4, no. 3, pp. 744–755, Sep. 2023.

[28] Z. Sun, M. Gao, M. Zhang, M. Lv, and G. Wang, "Research on recognition method of broiler overlapping sounds based on random forest and confidence interval," *Comput. Electron. Agricult.*, vol. 209, Jan. 2023, Art. no. 107801.
[29] I. Balabanova, S. Kostadinova, and G. Georgiev, "Stress recognition using sound analysis, k-NN, decision tree and artificial intelligence approach," in *Proc. Int. Conf. Biomed. Innov. Appl. (BIA)*, vol. 1, Jun. 2022, pp. 123–126.
[30] C. Pan, C. Shi, H. Mu, J. Li, and X. Gao, "EEG-based emotion recognition using logistic regression with Gaussian kernel and Laplacian prior and investigation of critical frequency bands," *Appl. Sci.*, vol. 10, no. 5, p. 1619, Feb. 2020.
[31] W. Zhao, Y. Lv, X. Guo, and J. Huo, "An investigation on early fault diagnosis based on naive Bayes model," in *Proc. 7th Int. Conf. Control Robot. Eng. (ICCRE)*, Apr. 2022, pp. 32–36.
[32] B. B. Hazarika, D. Gupta, and B. Kumar, "EEG signal classification using a novel Universum-based twin parametric-margin support vector machine," *Cognit. Comput.*, vol. 2023, pp. 1–16, Jan. 2023.
[33] V. Rupapara, F. Rustam, A. Amaar, P. B. Washington, E. Lee, and I. Ashraf, "Deepfake tweets classification using stacked bi-LSTM and words embedding," *PeerJ Comput. Sci.*, vol. 7, p. e745, Oct. 2021.
[34] J. Ding, Y. Wang, H. Si, S. Gao, and J. Xing, "Multimodal fusion-AdaBoost based activity recognition for smart home on WiFi platform," *IEEE Sensors J.*, vol. 22, no. 5, pp. 4661–4674, Mar. 2022.
[35] A. Raza, K. Munir, M. Almutairi, F. Younas, M. M. S. Fareed, and G. Ahmed, "A novel approach to classify telescopic sensors data using bidirectional-gated recurrent neural networks," *Appl. Sci.*, vol. 12, no. 20, p. 10268, Oct. 2022.
[36] E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Cham, Switzerland: Springer, 2019.
[37] J. Li, J. Guo, M. Ma, Y. Zeng, C. Li, and J. Xu, "A gunshot recognition method based on multi-scale spectrum shift module," *Electronics*, vol. 11, no. 23, p. 3859, Nov. 2022.

ALI RAZA received the B.Sc. degree in computer science from the Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim Yar Khan, Pakistan, in 2021, where he is currently pursuing the M.S. degree in computer science. His current research interests include data science, artificial intelligence, data mining, natural language processing, machine learning, deep learning, and image processing.

FURQAN RUSTAM received the M.C.S. degree from the Department of Computer Science, Islamia University of Bahawalpur, Pakistan, in October 2017, and the master's degree in computer science from the Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim Yar Khan, Pakistan. He is currently pursuing the Ph.D. degree in computer science with University College Dublin, Ireland. He was a Research Assistant with the Fareed Computing and Research Center, KFUEIT. His current research interests are related to data mining, machine learning, and artificial intelligence, mainly working on creative computing and supervised machine learning.

BHARGAV MALLAMPATI received the master's degree from the Department of Electrical Engineering, University of North Texas. His current research interests are related to data mining, machine learning, and artificial intelligence, mainly working on creative computing and supervised machine learning.

PRADEEP GALI received the master's degree from the Department of Electrical Engineering, University of North Texas. He is currently working as a Solution Engineer in Zoom Video Communications. His current research interests are related to data mining, machine learning, and artificial intelligence, mainly working on creative computing and supervised machine learning.

IMRAN ASHRAF received the M.S. degree in computer science from the Blekinge Institute of Technology, Karlskrona, Sweden, in 2010, and the Ph.D. degree in information and communication engineering from Yeungnam University, South Korea, in 2018. He was a Postdoctoral Fellow with Yeungnam University. He is currently an Assistant Professor with the Information and Communication Engineering Department, Yeungnam University, Gyeongsan, South Korea. His research areas include indoor positioning and localization in 5G and beyond, indoor location-based services in wireless communication, smart sensors for smart cars, and data analytics.