

AN AUTOMATED GUNSHOT AUDIO CLASSIFICATION METHOD BASED ON FINGER PATTERN FEATURE GENERATOR AND ITERATIVE RELIEFF FEATURE SELECTOR

Türker TUNCER¹, Şengül DOĞAN², Erhan AKBAL³, Emrah AYDEMİR^{4*}

¹⁻²⁻³Fırat University, Technology Faculty, Department of Digital Forensics Engineering, Elazığ, 23119, Turkey

²Sakarya University, Faculty of Management, Department of Management Information Systems, Sakarya, 54050, Turkey

Geliş tarihi: 05.05.2021 Kabul tarihi: 23.06.2021

ABSTRACT

Audio forensics applications and methods are very crucial to clarify crimes. To accelerate audio analysis process and classify audios with high accuracy, machine learning (ML) methods must be used in audio forensics. An automated gunshot audios classification method is presented in this study. To implement our automated gunshot classification method, a novel gun audios dataset was collected from YouTube with 8 classes in the first phase. A novel ML method is presented in the second phase and the proposed ML method contains three fundamental phases. These phases are a novel finger pattern (finger-pat), statistical moments and discrete wavelet transform (DWT) based feature generation network, informative/distinctive feature selection with iterative ReliefF (IRF) feature selector and classification with a k nearest neighbors (kNN) classifier (shallow) to show success of the generated and selected features by using the proposed finger-pat based feature generation network and IRF feature selector. These methods and kNN achieved 94.48% classification accuracy. These results demonstrate that our proposed method can be used in gunshot audio analysis.

Keywords: Gunshot audio classification; finger pattern; iterative ReliefF; audio forensics; machine learning

1. Introduction

Nowadays, digital evidences are crucial to elucidation for a crime case. Since, humans have been used electronic devices everywhere and these devices has memory [1, 2]. Valuable evidences have generally been acquired by examining these electronic devices memory and crime investigators have extracted knowledge from these digital data. This situation made it compulsory to use machine learning (ML) in the digital forensics to improve efficiency. Identification and classification of weapons according to gunshot audios are important in security, military and scientific research [3]. Therefore, setup scenarios are prepared and the data is obtained with microphones by repeating the gun shot. But the audios in the scene forensic applications are different from the setup scenario [4, 5]. Because the angle and position information of the weapon is not known. Determining the category, caliber and model of the weapon used at the scene helps the investigation process and suspect identification. Thus, it is necessary to detect independently of the physical conditions such as the shooter's position.

e-posta¹: turkertuncer@firat.edu.tr ORCID ID: <https://orcid.org/0000-0002-1425-4664>

e-posta²: sdogan@firat.edu.tr ORCID ID: <https://orcid.org/0000-0001-9677-5684>

e-posta³: erhanakbal@firat.edu.tr ORCID ID: <https://orcid.org/0000-0002-5257-7560>

* e-posta⁴: emrahaydemir@sakarya.edu.tr ORCID ID: <https://orcid.org/0000-0002-8380-7891> (Sorumlu Yazar)

Audio forensic is the sub-branches of digital forensic. Audio forensic is about collecting, analyzing and evaluating audio recordings that can be used as evidence by justice or law enforcement officers [6]. Investigation of crimes committed using firearms is the subject of the audio forensic [7]. The audios produced by gunshot at the scene contains important information about the caliber, model and category of the gun used by the criminal [8]. Gun shooting causes audio at the scene. The audio signal differs to the characteristics of the weapon. The suspect may escape from the scene or destroy the weapon. In this case, it can obtain information about the gun and the audios to be taken from the devices that record audio, such as security camera recordings and mobile phones [9]. Therefore, audio signal in setup scenarios cannot be used. Audio signals for variable condition are needed. In our study, it is aimed to determine the model, category and caliber from the gunshot in any position, angle and space under variable conditions. The methods working in variable conditions are also successful in setup scenarios.

1.1. Motivation

Gunshot audio forensic examiner provide a lot of information for law enforcement and investigators.

The main motivations of our study are given below;

- Gunshots with unknown source can be detected using audio signals, which will be collected at the scene. Audio data can be obtained such as security cameras, wearable devices, mobile phone videos, and mobile recorders may be the sources. Using these audio signals, we can learn the brand, category and caliber of the weapon.
- Provides information about the weapon if a gunshot audio is encountered in an audio file obtained from digital evidence.
- The gun cannot be clearly visible on the scene camera, or there may not be an empty shell case at the scene. Gun information can be learned if there is a gunshot audio in the camera recordings.

1.2. Related works

Recent year ambient audio/sound classification and detection are among the popular topics. Ambient sounds are classified for many purposes. There are many studies in different areas such as environmental monitoring [10-12], underwater monitoring [13, 14], health [15, 16], construction and design [17], animal tracking [18], IT crimes [19]. Audio Forensics, the sub-branch of digital forensic, is the science that studies criminal audios. Investigators obtain many audio files from the scene or storage devices. These files may contain files with gunshots. Information about the weapon is used for the detection of crime. Guns are evidence in many cases such as murder, injury, extortion. Therefore, files with gunshots should be examined carefully. Guns are generally classified according to caliber, brand, model and category (such as gun rifle) information from gunshots. There are several studies in the literature on the analysis of gunshot. The acoustic features of gunshot audios are different. Researchers have proposed various methods and algorithms for the classification of audio signals. Begault and Beck [8] presented forensic audio gunshot analysis methods. In their study, the limitations encountered in the detection of gun audios and the reliability of the techniques were shown. Busse et al. [20] presented a model for gunshot classification. It was determined whether there was gunshot in the audio signal. The support vector machine (SVM) classifier was used and achieved 70.39% accuracy. Khan et al. [21] used the hierarchical Gaussian mixture model (GMM) Classification method for gunshot detection. He achieved 90% category classification in a total of 100 shots data sets in 10 different gun types. Ahmed et al. [22] proposed a low-cost system for the recognition of gunshot. Mel-Frequency Cepstral Coefficients (MFCC) and SVM classifier were used in their study. Djeddou and Touhami [23] aimed to classify the gunshot audio obtained from a noisy environment. GMM classifier using cepstral features was used. Audio signals obtained from the barrel mouth of 5 different guns are classified with a rate of 96.29%. Kiktova et al. [24] proposed a method for determining whether there is a gunshot in the audio

signals obtained from noise monitoring stations in cities. Hidden Markov Model (HMM) classification and viterbi based decoding algorithm obtained 80% successful detection rate. Khan et al. [25] identified gun types using the exemplar embedding hierarchical GMM classification method. Their datasets consist of total of 100 gunshots by 20 different gun types. An accuracy rate of 60-72% was obtained. Morton Jr et al. [26] provided a method for classifying gunshot audio signals. These signals are modeled using HMM with autoregressive (AR). Raponi and Ali [7] presented a method for classifying gunshots using CNN. The dataset consists of 3655 audios of 59 different weapons and achieved 90% accuracy classification rate. Sánchez-Hevia et al. [27] presented a model using LS-LDA and maximum likelihood decision fusion method. The dataset contains 840 gunshot signals of 14 gun types. Classification rate of 94.1% was obtained by using cepstral, spectral and temporal features.

1.3. Our Method

The proposed method aimed to achieve high classification accuracy like other ML methods. ML methods generally consists of three main phases and these are feature generation, feature selection and classification. Appropriate methods should be presented/used for these phases. Generally, existing powerful methods have been used to guarantee high success rates. Novel methods are risky methods. However, we preferred propose novel method in this journal. In the feature generation phase, low, medium and high levels features should be extracted and these features can be extracted using multileveled or multilayered feature generation networks such as deep learning network. Therefore, 12 levelled (The number of levels are determined as experiments. Trail and errors method was used in this phase) a novel feature generation network is presented. Local feature generators such as ternary pattern, local binary pattern which are the mostly known descriptors aim generate global optima features using local relationships and kernels. They have low computational complexity, separable feature generation ability and clear mathematical background. Therefore, they have been widely used in the literature. Since, a novel nature inspired (finger-pat) one dimensional local feature generator is presented to extract textural features of the gunshot audios. Statistical features are also extracted of both textural features and raw signal.

To select informative feature of the extracted features by a multileveled feature generation network, an effective feature selector should be used. Therefore, we used IRF feature selector to solve the automated best feature vector selection problem.

Any classifier can be used in the classification phase since the effective feature are generated and selected. Therefore, kNN [28] classifier is selected to show results.

1.4. Contributions

This work proposes a novel gun audio classification method and our contributions are;

- Gun audio classification is one of the most crucial for crime investigator. Therefore, an automated gun audio classification method must be presented for the audio forensics field. However, testbed is the most problem for developing gun audio classification method. By using automated gun audio classification methods, novel intelligent audio forensics toolboxes can be developed. To solve testbed problem, a novel gun dataset was collected. This dataset contains 851 audios of the 8 gun classes.
- A hand-crafted and nature inspired method is presented. The used three fundamental components/techniques of this method are finger-pat and statistical features based feature generation network, IRF feature selector and kNN classifier. A new finger-pat which is a nature inspired feature extractor is presented in this work. This method is a nature inspired feature extractor to obtain textural features. In the feature generation network both textural features by using the proposed finger-pat and statistical feature with 19 widely preferred

statistical moments are extracted. Also, features of feature are generated in this work. Since, statistical features of the finger-pat features are generated in our network. An exhaustive and informative feature generation is proposed.

- ReliefF [29] cannot select the best number of features automatically. To solve this problem and obtain the best feature vector, IRF method is utilized as feature selector.
- The proposed finger-pat based feature generation network and IRF feature selector extract comprehensive and select informative feature. To specify this situation, kNN classifier is used and 94.48% classification accuracy was calculated by using kNN with 10-fold cross validation.

2. Audio dataset

Varying environment model was used to gather gunshot audios since YouTube was used to data source. The audios of the gun models were collected on YouTube [30] using videos that are open to everyone. These files gathered on different dates were downloaded and parts were produced for 2 seconds for each gun type. In this way, a total of 851 files types were obtained with 8 gun models. The sampling rate of the audio files is 44100 Hz. Firstly, each audio file was converted to wav file format. And then, the audios were carefully listened to and made sure that there were no different audios or noises inside. It has been carefully checked that the same audio does not continue repeatedly during the fragmentation of the audio files. WavePad Audio Editor program was used for all these processes. The audio files are stored at <https://www.kaggle.com/emrahaydemr/gunshot-audio-dataset> for using by other researchers. Table 1 denotes details of the number of audios formed are listed.

Table 1. Details of the collected gunshots audio dataset

ID	Model	Number of observations
1	AK-47	72
2	IMI Desert Eagle (Desert Eagle)	100
3	AK-12	98
4	M16	200
5	M249	99
6	MG-42	100
7	MP5	100
8	Zastava M92	82

2.1. The proposed finger pattern and iterative ReliefF based gunshot audio classification method

We present a novel audio classification method for gunshots and this method consist of three fundamental phases. These are finger-pat and statistical moments based feature generation network, feature selection with IRF feature selector and classification with kNN [28]. The proposed finger-pat and IRF based audio classification method aims achieve high classification ability using conventional classifiers like kNN. Graphical overview of the proposed finger-pat and IRF based audio classification method is shown in Fig. 1.

To overview the proposed finger-pat and IRF based classification method, procedure of this method is shown in Fig. 2.

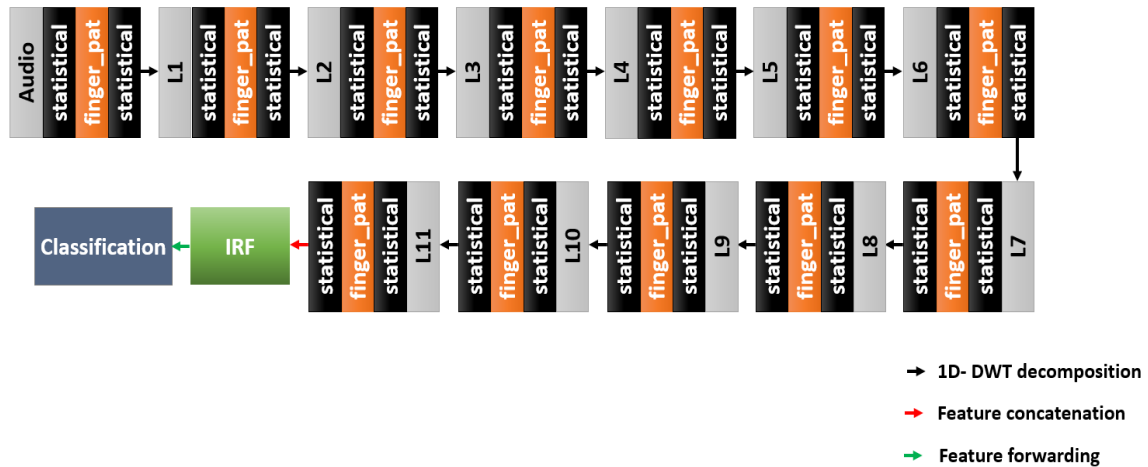


Fig. 1. Graphical summarization of the proposed finger-pat and IRF based method.

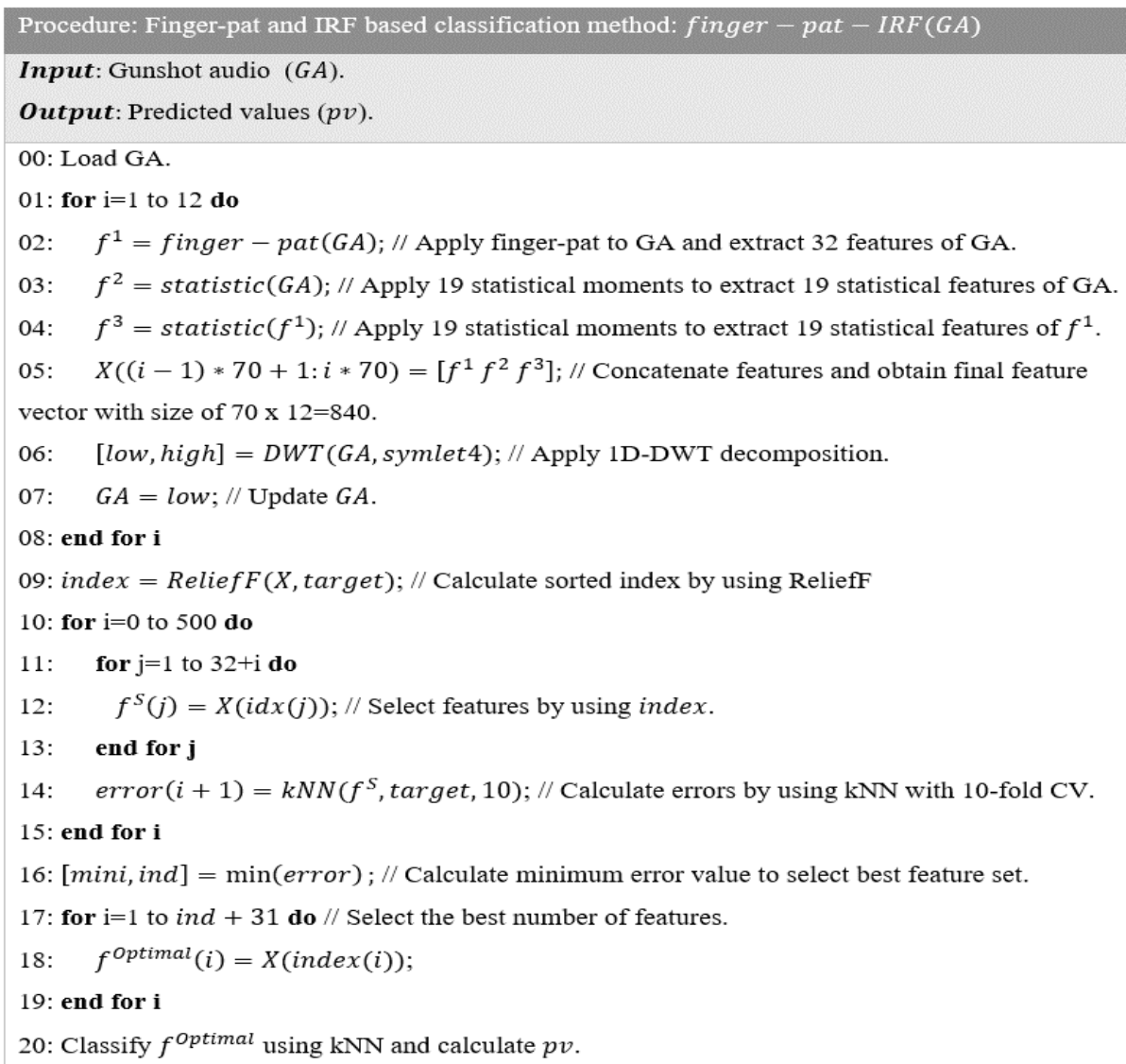


Fig. 2. Procedure of the proposed finger-pat and IRF based audio classification method.

Fig. 2 indicates pseudo code of our method. Lines 01-08 denote finger-pat and statistical moments based multileveled generation network, lines 09-19 show IRF feature selector based informative feature selection process and line 20 demonstrates classification phase. Details of the proposed finger-pat and IRF based classification method are given in subsections.

2.2. Feature generation network

This work proposes a novel feature generation network and this network aims to generate high, medium and low levels features. Therefore, multilevel DWT [31, 32] transform is utilized as level creation method. In each level, finger-pat and 19 statistical moments generate 70 features. The interesting side of this network is to extract features of features. The proposed feature generation network extracts both textural and statistical features together as shown Fig. 3.

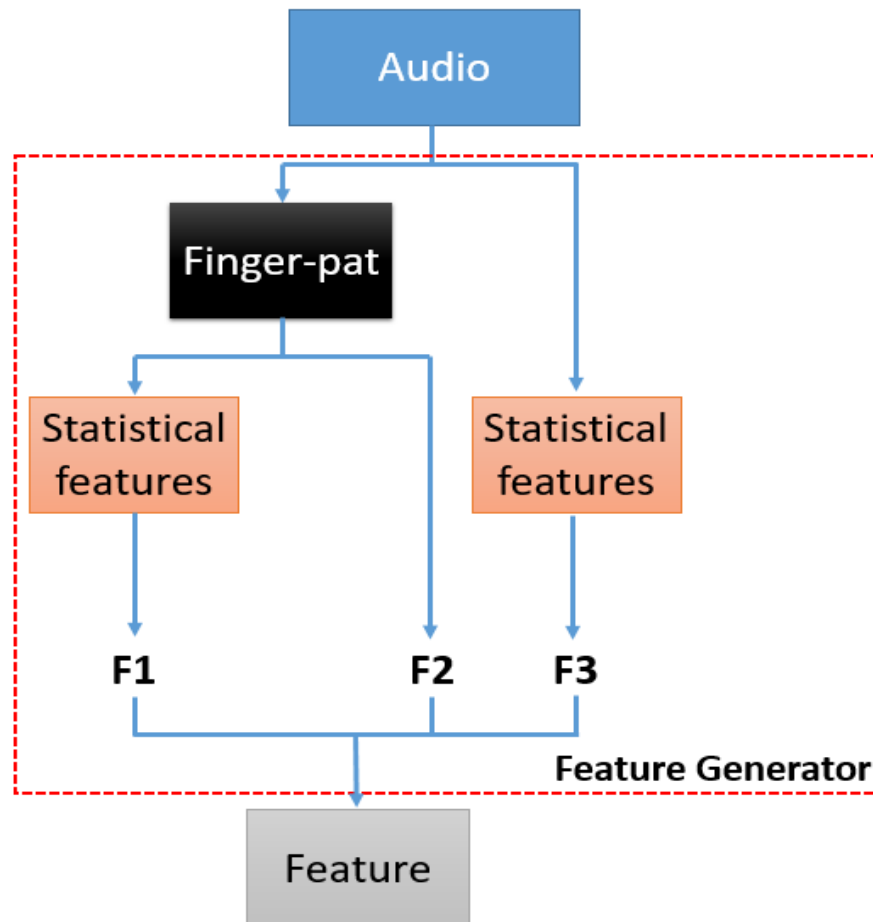


Fig. 3. Graphical structure of the proposed finger-pat and statistical moments based feature generator.

Steps of the proposed finger-pat and statistical moments based multilevel feature generation network are;

Step 1: Generate 11 low-pass filter sub-bands by using multilevel DWT with symlets 4 filter (*sym4*). Symlets 4 filter is very effective for both signal decomposition and signal denoising.

$$[LowPass^1, HighPass^1] = DWT(GA, sym4) \quad (1)$$

$$[LowPass^k, HighPass^k] = DWT(LowPass^{k-1}, sym4), k = \{2, 3, \dots, 11\} \quad (2)$$

where $LowPass^k$ and $HighPass^k$ are k^{th} low pass and high pass filters of the DWT.

Step 2: Generate 32 features from the raw audio and low pass filter sub-bands of it by using fingerprint and extract 19 features from each 32 features using statistical feature generation function (*sfgf*). Mathematical explanation of this functions are shown as below.

$$s1 = \frac{\sum_{i=1}^L signal_i}{L} \quad (3)$$

$$s2 = \sqrt{\frac{\sum_{i=1}^L (signal_i - s1)^2}{L}} \quad (4)$$

$$s3 = \sqrt{\frac{\sum_{i=1}^L (signal_i)^2}{L}} \quad (5)$$

$$s4 = \frac{\sum_{i=1}^L signal_i}{s3} \log\left(\frac{signal_i}{s3}\right) \quad (6)$$

$$s5 = \sum_{i=1}^L \frac{i * signal_i - f1}{\sigma_i} \quad (7)$$

$$s6 = \frac{\sum_{i=1}^L |signal_{i+1} - signal_i|}{L} \quad (8)$$

$$s7 = \frac{\sqrt{L(L-1)}}{L-2} \left(\frac{\frac{1}{L} \sum_{i=1}^L (signal_i - s1)^3}{\frac{1}{L} \sum_{i=1}^L (signal_i - s1)^2} \right) \quad (9)$$

$$s8 = \frac{L-1}{(L-2)(L-3)} \left[(L+1) \left(\left(\frac{\frac{1}{L} \sum_{i=1}^L (signal_i - s1)^4}{\frac{1}{L} \sum_{i=1}^L (signal_i - s1)^2} \right) - 3 \right) + 6 \right] \quad (10)$$

$$s9 = signal\left(\frac{L}{2}\right) \quad (11)$$

$$s10 = \{signal\} \quad (12)$$

$$s11 = \max\{signal\} \quad (13)$$

$$s12 = \frac{s1}{s2} \quad (14)$$

$$s13 = \sqrt{\frac{\sum_{i=1}^L signal_i^2}{L}} \tag{15}$$

$$s13 = \frac{\sqrt{\frac{1}{L} \sum_{i=1}^L signal_i^2}}{\frac{1}{L} \sum_{i=1}^L |signal_i|} \tag{16}$$

$$s15 = \frac{\max |signal|}{\sqrt{\frac{1}{L} \sum_{i=1}^L signal_i^2}} \tag{17}$$

$$s16 = \frac{\max |signal|}{\frac{1}{L} \sum_{i=1}^L \sqrt{|signal_i|^2}} \tag{18}$$

$$s17 = \frac{\max |signal|}{\frac{1}{L} \sum_{i=1}^L |signal_i|} \tag{19}$$

$$s18 = (\max(signal) - \min(signal)) \tag{20}$$

$$s19 = (\max(signal) - \text{mean}(signal)) \tag{21}$$

Where signal is input audio signal or low pass filter of it. s defines statistical moment, L represents length of the signal. These equations create *sfgf*. Feature generation method of this step is shown in below.

$$f^1(1:32) = \text{finger} - \text{pat}(GA) \tag{22}$$

$$f^3(1:19) = \text{sfgf}(f^1(1:32)) \tag{23}$$

$$f^1(k * 32 + 1:(k + 1) * 32) = \text{finger} - \text{pat}(\text{LowPass}^k), k = \{1,2, \dots, 11\} \tag{24}$$

$$f^3(k * 19 + 1:(k + 1) * 19) = \text{sfgf}(f^1(k * 32 + 1:(k + 1) * 32)) \tag{25}$$

As can be seen from Eqs. 22-25, *finger - pat(.)* represents as feature generation function of the finger-pat. Eqs. 22-25 define both feature generation and fusion.

Humans have 10 finger on their hands. When the hands are put together, it is seen that the fingers arrangement has a center symmetric structure as shown in Fig. 4.



Fig. 4. Arrangement of the hand fingers.

Therefore, a 10 sized overlapping block is used and center symmetric pattern is used for feature generation. Steps of the proposed finger-pat are shown as below.

Step 2.1: Divide signal into 10 sized overlapping blocks.

$$p_h = \text{signal}(i:i+9), h = \{1,2, \dots, L-9\} \quad (26)$$

where L is length of signal.

Step 2.2: Generate binary features by using center symmetric pattern and signum function.

$$\text{bit}(y) = \text{signum}(p(y), p(11-y)), y = \{1,2, \dots, 5\} \quad (27)$$

$$\text{signum}(fp, sp) = \begin{cases} 0, & fp - sp < 0 \\ 1, & fp - sp \geq 0 \end{cases} \quad (28)$$

where fp and sp denote first and second parameters of the signum function respectively.

Step 2.3: Create map value of the finger-pat by using Eq. 29.

$$\text{map}(i) = \sum_{k=1}^5 \text{bit}(k) * 2^{k-1} \quad (29)$$

Step 2.4: Calculate histogram of the map value.

$$\text{histo}(r) = 0, r = \{1,2, \dots, 2^5\} \quad (30)$$

$$\text{histo}(\text{map}(i) + 1) = \text{histo}(\text{map}(i) + 1) + 1 \quad (31)$$

In Eq. 30, initial values of the histogram (histo) are assigned as 0 and histogram extraction is processed in Eq. 31.

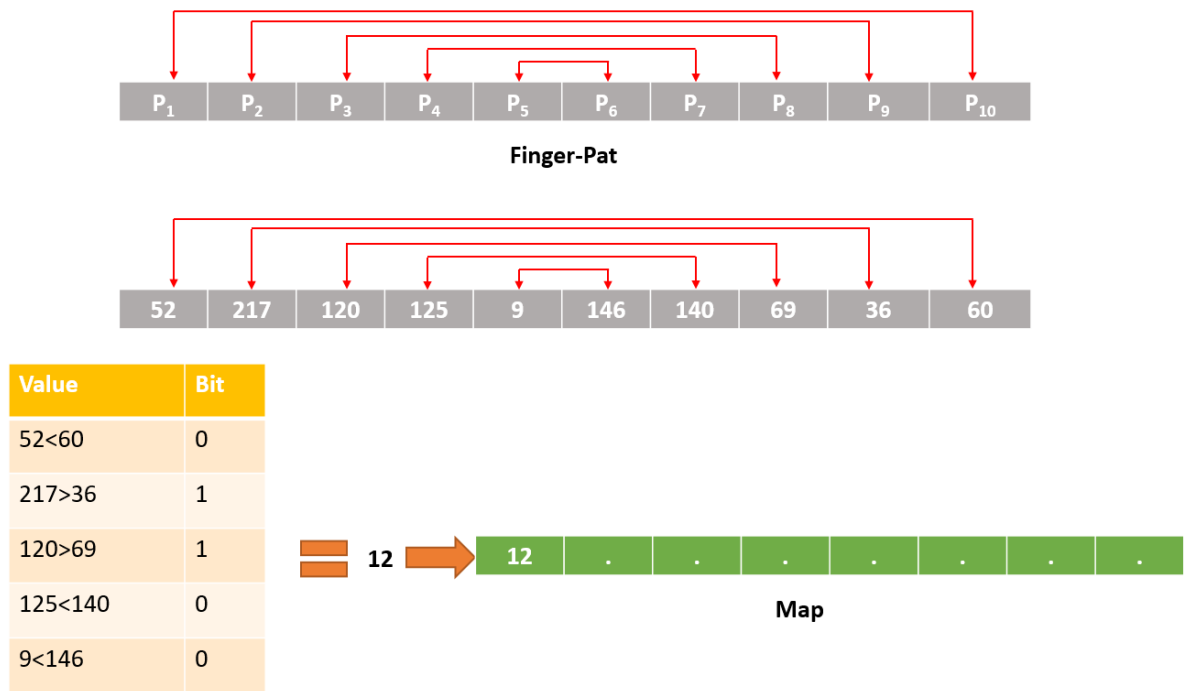


Fig. 5. Pattern of the proposed finger-pat and a numerical example about finger-pat. Histogram of the calculated map is utilized as feature vector. Thus, map value calculation is shown in this figure by using a numerical example.

Step 3: Extract 19 statistical features of the raw gunshot audio and low pass filter of it. Eqs. 32 and 33 define feature generation and concatenation.

$$f^2(1:19) = sfgf(GA) \tag{32}$$

$$f^2(k * 19 + 1:(k + 1) * 19) = sfgf(LowPass^k), k = \{1,2, \dots,11\} \tag{33}$$

As seen Steps 1-3, three type features are extracted and these are called as f^1 (finger-pat features, textural features), f^2 (statistical features) and f^3 (features of features, statistical features of the finger-pat features). Sizes of these features vectors are 384, 158 and 158 respectively. These features are concatenated and final feature vector (X) with size of 840 is obtained.

Step 4: Combine generated three feature vectors (f^1, f^2, f^3) to generate X . Mathematical definition of the feature concatenation is shown in Eqs. 34-36.

$$X(i) = f^1(i), i = \{1,2, \dots,384\} \tag{34}$$

$$X(j + 384) = f^2(j), j = \{1,2, \dots,158\} \tag{35}$$

$$X(j + 542) = f^3(j), j = \{1,2, \dots,158\} \tag{36}$$

2.3. IRF based feature selection

ReliefF [29] is one of the mostly preferred feature selector for informative feature selection and generates weights which are both positive and negative for each feature. By using the generated feature weights, feature selection is processed in ReliefF. Negative weighted features can be assigned as redundant features in the ReliefF. However, there is no automatic optimal number of features selection method. To solve this problem, a new generation ReliefF is presented and it is named as IRF. IRF uses ReliefF weights to select optimal features. In the IRF, a loss value generator, we selected classifiers as loss value generator, and an iteration are used to select optimal features. Steps of our presented feature selector and graphical summarization of the IRF are shown in below.

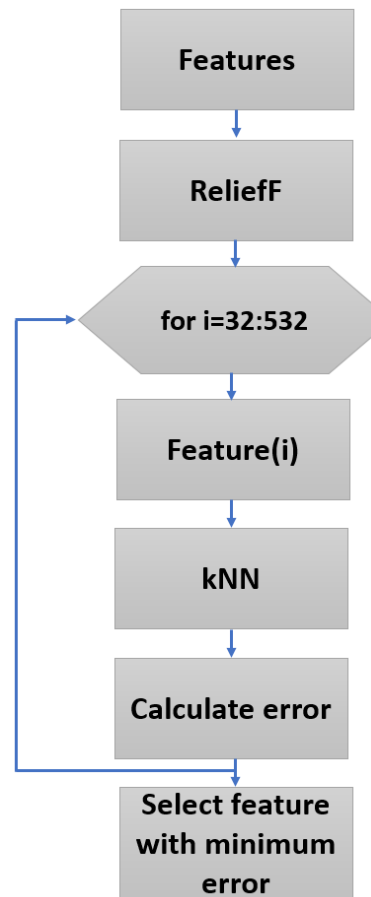


Fig. 6. Flow diagram of the used IRF feature selector. The proposed methods select from 32 features to 532 features and it calculates error rate of each features by using kNN classifier.

Step 5: Generate ReliefF weights of the X .

$$weight = ReliefF(X, target) \quad (37)$$

Step 6: Calculate index values of the sorted features.

$$[maxi, index] = sort(weight) \quad (38)$$

where $sort(.)$ is sorting function.

Step 7: Select features. IRF has high computational complexity, hence, 501 features are selected in this work to decrease complexity.

$$feat^{S-31}(i) = X(index(i)), i = \{1, 2, \dots, S\}, S = \{32, 33, \dots, 532\} \quad (39)$$

Step 8: Calculate loss values of the selected features by using kNN with 10-fold CV.

$$loss(S - 31) = kNN(feat^{S-31}, target, 10) \quad (40)$$

Step 9: Find index (*ind*) of the minimum loss value.

$$[mini, ind] = \min(loss) \quad (41)$$

$$ind = ind + 31 \quad (42)$$

The proposed IRF starts selecting features from 32 features. Therefore, Eq. 42 is used to calculate optimal number of features.

Step 10: Select optimal number of features.

$$f^{Optimal}(i) = X(index(i)), i = \{1, 2, \dots, ind\} \quad (43)$$

where $f^{Optimal}$ is optimal features.

2.4. Classification

The last phase of our method is classification. kNN classifier is used in this work. Parameters of the kNN are given as follows. k value is selected as 1 and distance metric is chosen Manhattan (City Block) distance metric. Mathematical explanation of the Manhattan distances is shown as below.

$$dist(A, B) = \sum_{i=1}^D |A_i - B_i| \quad (44)$$

This kNN classifier is used for both feature selection (loss function of the IRF feature selector) and classification. The last step (Step 11) is shown as below.

Step 11: Classify $f^{Optimal}$ by using kNN.

$$pv = kNN(f^{Optimal}, target, 10) \quad (45)$$

2.5. Experiments

We collected a novel dataset from YouTube [30]. Therefore, these gunshot audios were collected from varying environments. We did not use any experimental setup for gunshot audios recording. A

novel ML method was presented to classify these gun-shots with high accuracy. Our experimental setup for designing the proposed finger-pat and IRF based classification method is given as follows. The proposed finger-pat and IRF based audio classification was coded in MATLAB (2018b) programming environment on a desktop computer (DC). Our DC has 32 GB RAM, 3.2 GHz 7th generation i7 CPU and Windows 10.1 operating system.

To evaluate performance of the used classifier (kNN). Accuracy (*cac*), unweighted average precision, recall, F1 (*UAR*, *UAP*, *F1*) and geometric mean (*gm*) are utilized [33-35]. We used multiclass classification. Therefore, these performance metrics have been used. Mathematics of them;

$$UAP = \frac{1}{T} \sum_{i=1}^T \frac{TP_i}{TP_i + FP_i} \quad (46)$$

$$UAR = \frac{1}{T} \sum_{i=1}^N \frac{TP_i}{TP_i + FN_i} \quad (47)$$

$$gm = \sqrt[T]{\prod_{i=1}^T \frac{TP_i}{TP_i + FP_i}} \quad (48)$$

$$F1 = \frac{2 * UAP * UAR}{UAP + UAR} \quad (49)$$

$$cac = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \quad (50)$$

where *TP*, *TN*, *FP*, *FN* and *T* are number of true positives, number of true negatives, number of false positives, number of false negatives and number of classes. The obtained measurements were listed in Table 2.

Table 2. The calculated performance measurements by using kNN.

Measurement	Rate (%)
Classification accuracy	94.48
Unweighted Average Recall	93.92
Unweighted Average Precision	94.91
F1-score	94.41
Geometric Mean	93.72

To validate these results, confusion matrix of this method is shown in Fig. 7.

		Predicted Class							
		1	2	3	4	5	6	7	8
True Class	1	68	1	1		1		1	
	2		86	1	7	4	2		
	3			98					
	4				200				
	5		4		1	94			
	6		4		2	1	93		
	7	1	3	1	3	5	3	83	1
	8								82

Fig. 7. The confusion matrix resulted by using kNN classifier.

As seen from Fig. 7, perfect classification rate (100.0% classification accuracy) was founded in three classes which are 3rd, 4th and 8th classes.

3. Discussions

For the gunshot analysis, ballistic and chemometrics analysis have generally used in crime scene. Manual audio analysis for gunshots are very difficult and a lot of time is needed to analyze these audios. Nowadays, time is one of the most valuable thing. Audio forensics analysis are processed manually by an expert. Therefore, these processes are depended on the audio forensic expert. Success of case is directly proportional success of the crime investigator (audio forensic analysis expert). To decrease error rate of this processes and eliminate waste of time, ML based automated audio classification methods should be used. This paper is proposed finger-pat and IRF based gunshot audio classification method to automatically analysis gunshots in crime investigation. This finger-pat and IRF based classification method is a cognitive method because problems of the ML are solved in this method step by step. This method is a nature inspired ML method. By using the proposed cognitive ML methods and other ML methods, intelligent audio analysis agents/toolboxes can be developed. According to audio forensics examiners, the proposed method will save a lot of time for audio analyses and it will not be necessary to use chemometrics and ballistic analysis for some cases.

We presented a new generation ML method. This method uses a finger-pat based feature generation network. The interesting side of the proposed feature generation network is to extract features of features. The proposed new generation gunshot audio classification method has a multileveled feature generation network. This network generates 840 features and IRF selects 517 the most informative features. Plotting of the IRF based feature selection process is shown in Fig. 8.

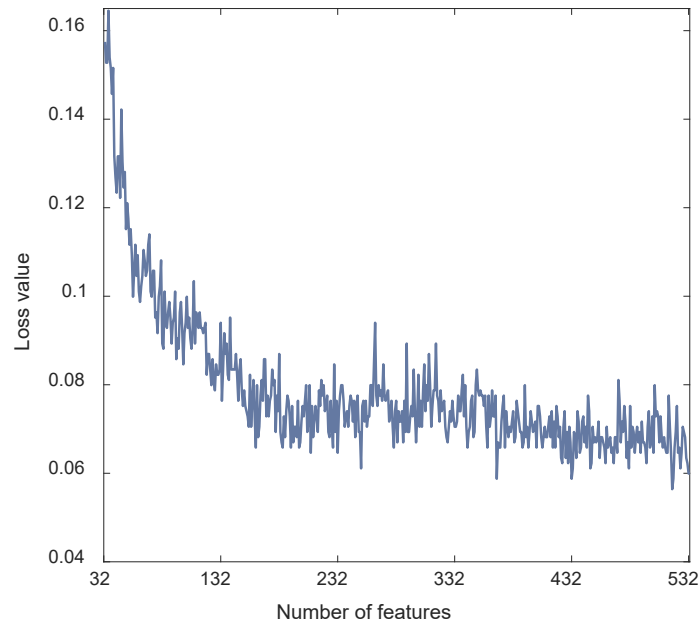


Fig. 8. Feature selection process of the IRF feature selector.

Fig. 8 shown error rates of each selected features. To clearly indicate success of the proposed finger-pat and IRF based gunshot audio classification method, this method was compared to other gunshot classification methods and results were listed in Table 3.

Table 3. Results of the proposed method and other gunshot classification methods.

Study and Year	Dataset	Method and Classifier	Accuracy Rate	Signal Acquisition Conditions	
				Varying	Setup scenario
[21]	10 gun and 100 gunshot	Gaussian Mixture Model (GMM)	90% category	-	+
[25]	20 gun and 100 gunshots	Exemplar embedding using hierarchical GMM	95-100% (category) 60-72% (model)	+	-
[36]	4 gun and 194 gunshots	Time-of-arrival and direction-of-arrival estimation	86% (caliber)	-	-
[26]	5 gun and 46 gunshots	Hidden Markov model (HMM)	95.65% (model)	-	+
[23]	5 gun and 230 gunshots	Hierarchical GMM classification	96.29% (category)	-	+
[24]	4 gun and 372 gunshot	HMM	80% (model)	-	+
[27]	14 gun and 840 gunshots	LS-LDA & Maximum Likelihood decision fusion	94.1% (model)	+	-

[20]	1 gun 374 gunshots	SVM	50% (Binary classification)	+	-
[7]	59 gun models 3655 gunshots	Convolutional Neural Networks	90% (Category, Caliber, and Model)	+	+
Our Proposed Method	8 gun models with 851 gunshots	Finger Pattern Feature Generator and Iterative ReliefF Feature Selector	94.48% (model) classification accuracy	+	+

As can be seen Table 3, the proposed achieved high classification accuracy for detecting model of the used guns. Also, the gathered dataset has 851 audios and it is a heterogeneous dataset. In this view, it is a good testbed to evaluate general success of the classification methods. This dataset clearly demonstrated of general success of the finger-pat and IRF based method. Benefits of our methods are;

- There are limited gunshot datasets in the literature. Therefore, a new gunshot dataset was collected and it was publicly published in this work.
- A new nature inspired feature extractor is presented and it is called as finger-pat. We used finger-pat as one of the fundamental feature extractors of our method.
- The optimal number of features are selected by using IRF. This result denoted that IRF solves optimal number of features selection problem.
- This paper presented a new gunshot audios classification method by using finger-pat and statistical moments based feature generation network and IRF feature selector together. The selected features were forwarded to kNN and 94.48% classification was achieved. This situation demonstrated success of the finger-pat and IRF based feature generation and selection methods.
- Robust classification results were obtained. Since, 10-fold CV was used as validation and test strategy.
- A high accurate method is presented. The proposed finger-pat and IRF based method resulted high accuracy on a heterogonous dataset.
- This method uses hand-crafted features and these methods are both simply and distinctive. They also have low computational complexity. In the proposed finger-pat and IRF based method, there is no need to set millions of parameters as in deep learning networks. Therefore, this method is a lightweight method.
- This paper defined ML feature generation and feature selection problems and it solved them step by step. In this view, this method is a cognitive method.
- There are limited ML based applications in the audio forensics. This work aimed solve this problem because ML based applications save a lot of time in the audio forensics examinations according to audio forensics experts.

4. Conclusion

Audio forensics is one of the mostly crucial sub-branches of the digital forensics. However, mostly audio forensics cases have been analyzed manually. Expertise level of audio forensics examiner directly affects to case. In order to minimize human mistakes, ML based methods must be used. Gunshot audio classification is one of the hard problems of the audio forensics. Therefore, we presented a novel automated gunshot audio classification method. In the first phase, we collected a new gunshot audios dataset from YouTube as a testbed. A new generation textural feature extraction function is proposed in this work and this feature extractor is named as finger-pat. The presented finger-pat was inspired by arrangement of human hands fingers. To generate comprehensively and multileveled features, 19 statistical moments, finger-pat and DWT were used together. IRF is a new generation variation of the ReliefF and it solved optimal features selection problem of the ReliefF feature selector. The selected most informative features were classified by using kNN. kNN reached 94.48% classification accuracy for gun models classification. The performance of the developed finger-pat and IRF based gunshot classification model was assessed by audio forensics experts and they confirmed effectiveness of this model.

5. Future Directions

Our future directions are;

- Gunshot audio classification is a hard problem for audio forensics experts. Therefore, a novel ML based assistant systems can be developed by using the proposed finger-pat and IRF based method for gunshot audios classification.
- Chemometrics and ballistic analyzes are high costly. To decrease cost for gun model detection, novel ML based applications/models can be developed.
- The proposed finger-pat and IRF based feature generation network and feature selection model can be applied to other signals for classifications.
- Finger-pat is a nature inspired feature extractor. This work shows that a high accurate classification method can be developed. Many nature-inspired feature extractors can be proposed.
- We collected a new gunshot audios dataset. Bigger datasets can be collected and our model can be tested on bigger datasets.

References

- [1] Casey E. Digital evidence and computer crime: Forensic science, computers, and the internet: Academic press; 2011.
- [2] Turner P. Unification of digital evidence from disparate sources (digital evidence bags). Digital Investigation. 2005;2(3):223-228.
- [3] Case A, Cristina A, Marziale L, Richard GG, Roussev V. FACE: Automated digital evidence discovery and correlation. digital investigation. 2008;5:S65-S75.
- [4] Kenneally EE, Brown CL. Risk sensitive digital evidence collection. Digital Investigation. 2005;2(2):101-119.
- [5] Turner P. Selective and intelligent imaging using digital evidence bags. digital investigation. 2006;3:59-64.
- [6] Maher RC. Audio forensic examination. IEEE Signal Processing Magazine. 2009;26(2):84-94.
- [7] Raponi S, Ali I, Oligeri G. Sound of Guns: Digital Forensics of Gun Audio Samples meets Artificial Intelligence. arXiv preprint arXiv:2004.07948. 2020.
- [8] Begault DR, Beck SD, Maher RC. Overview of forensic audio gunshot analysis techniques. Paper presented at: Audio Engineering Society Conference: 2019 AES International Conference on Audio Forensics, 2019.

- [9] Maher RC. Acoustical characterization of gunshots. Paper presented at: 2007 IEEE Workshop on Signal Processing Applications for Public Security and Forensics, 2007.
- [10] Sueur J, Farina A. Ecoacoustics: the ecological investigation and interpretation of environmental sound. *Biosemiotics*. 2015;8(3):493-502.
- [11] Xie J, Zhu M. Handcrafted features and late fusion with deep learning for bird sound classification. *Ecological Informatics*. 2019;52:74-81.
- [12] Green M, Murphy D. Environmental sound monitoring using machine learning on mobile devices. *Applied Acoustics*. 2020;159:107041.
- [13] Mayer P, Magno M, Benini L. Self-sustaining acoustic sensor with programmable pattern recognition for underwater monitoring. *IEEE Transactions on Instrumentation and Measurement*. 2019;68(7):2346-2355.
- [14] Pham G-T, Baron V, Finez A, Nicolas B. High resolution source localization in underwater acoustics for deep sea mining monitoring. Paper presented at: OCEANS 2019-Marseille, 2019.
- [15] Lim SJ, Jang SJ, Lim JY, Ko JH. Classification of snoring sound based on a recurrent neural network. *Expert Systems With Applications*. 2019;123:237-245.
- [16] Patil S, Saxena A, Talreja T, Bhatti V. Medical Diagnosis of Ailments Through Supervised Learning Techniques on Sounds of the Heart and Lungs. *Soft Computing and Signal Processing*: Springer; 2019:253-262.
- [17] Jung M, Chi S. Human activity classification based on sound recognition and residual convolutional neural network. *Automation in Construction*. 2020;114:103177.
- [18] González-Hernández FR, Sánchez-Fernández LP, Suárez-Guerra S, Sánchez-Pérez LA. Marine mammal sound classification based on a parallel recognition model and octave analysis. *Applied Acoustics*. 2017;119:17-28.
- [19] Muhammad G, Alghathbar K. Environment recognition for digital audio forensics using MPEG-7 and mel cepstral features. *Journal of Electrical Engineering*. 2011;62(4):199-205.
- [20] Busse C, Krause T, Ostermann J, Bitzer J. Improved Gunshot Classification by Using Artificial Data. Paper presented at: Audio Engineering Society Conference: 2019 AES International Conference on Audio Forensics, 2019.
- [21] Khan S, Divakaran A, Sawhney HS. Weapon identification using hierarchical classification of acoustic signatures. Paper presented at: Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense VIII, 2009.
- [22] Ahmed T, Uppal M, Muhammad A. Improving efficiency and reliability of gunshot detection systems. Paper presented at: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013.
- [23] Djeddou M, Touhami T. Classification and modeling of acoustic gunshot signatures. *Arabian journal for science and engineering*. 2013;38(12):3399-3406.
- [24] Kiktova E, Lojka M, Pleva M, Juhar J, Cizmar A. Gun type recognition from gunshot audio recordings. Paper presented at: 3rd International Workshop on Biometrics and Forensics (IWBF 2015), 2015.
- [25] Khan S, Divakaran A, Sawhney HS. Weapon identification across varying acoustic conditions using an exemplar embedding approach. Paper presented at: Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IX, 2010.
- [26] Morton Jr KD, Torrione PA, Collins L. Classification of acoustic gunshot signatures using a nonparametric Bayesian signal model. Paper presented at: Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense X, 2011.
- [27] Sánchez-Hevia HA, Ayllón D, Gil-Pita R, Rosa-Zurera M. Maximum likelihood decision fusion for weapon classification in wireless acoustic sensor networks. *IEEE/ACM transactions on audio, speech, and language processing*. 2017;25(6):1172-1182.

- [28] Liao Y, Vemuri VR. Use of k-nearest neighbor classifier for intrusion detection. *Computers & security*. 2002;21(5):439-448.
- [29] Robnik-Šikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. *Machine learning*. 2003;53(1-2):23-69.
- [30] Youtube. <https://www.youtube.com/>. 2020.
- [31] Shensa MJ. The discrete wavelet transform: wedding the a trous and Mallat algorithms. *IEEE Transactions on signal processing*. 1992;40(10):2464-2482.
- [32] Jensen A, la Cour-Harbo A. *Ripples in mathematics: the discrete wavelet transform*: Springer Science & Business Media; 2001.
- [33] Schiele J, Rabe F, Schmitt M, et al. Automated Classification of Airborne Pollen using Neural Networks. Paper presented at: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019.
- [34] Stuhlsatz A, Meyer C, Eyben F, Zielke T, Meier G, Schuller B. Deep neural networks for acoustic emotion recognition: Raising the benchmarks. Paper presented at: 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP), 2011.
- [35] Tuncer T, Dogan S, Pławiak P, Acharya UR. Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals. *Knowledge-Based Systems*. 2019;186:104923.
- [36] Sallai J, Hedgecock W, Volgyesi P, Nadas A, Balogh G, Ledeczki A. Weapon classification and shooter localization using distributed multichannel acoustic sensors. *Journal of systems architecture*. 2011;57(10):869-885.