Applied Acoustics 177 (2021) 107916

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

A new fractal H-tree pattern based gun model identification method using gunshot audios

Sengul Dogan

Department of Digital Forensics Engineering, College of Technology, Firat University, Elazig Turkey

ARTICLE INFO

Article history:
Received 15 October 2020
Received in revised form 16 December 2020
Accepted 6 January 2021
Available online 27 January 2021

Keywords:
H-tree pattern
Audio forensics
Gun model identification

ABSTRACT

Background: Gun model identification (GMI) is a complex issue for digital forensics examiners/professions. Because the GMI process is a highly costed process, and it is generally detected manually. A sound classification model is presented in this research to decrease the cost of the GMI and automate this process.

Material and method: The primary objective of this research is to present a new intelligent audio forensics tool. Therefore, a new gunshot dataset was collected, and the collected dataset includes 2130 audios of the 28 gun models. This dataset can be downloaded using http://web.firat.edu.tr/sdogan/Gun_S_Dogan.rar link. The presented fractal H-tree pattern-based classification method is applied to these audios to obtain results. This method has three fundamental phases, and these are feature extraction, the most informative features selection, and classification. This method uses both a fractal textural generator and statistical features. By deploying tunable q-factor wavelet transform (TQWT), a multileveled feature generation method is created to generate both low-level and high-level features. The recommended fractal H-tree pattern and statistical feature extraction functions generate features at each level. Neighborhood component analysis (NCA) chooses the most informative features. In the classification phase, the support vector machine (SVM) and k nearest neighbor (kNN) classifiers are used.

Results: The recommended fractal H-tree pattern-based method yielded 96.10% and 90.40% by employing kNN and SVM, respectively.

Conclusion: The calculated results and findings denoted the high classification capability of the presented fractal H-tree pattern-based method for gun model classification using gunshot audios. Also, this research shows that a new audio forensic tool can be developed by employing the presented method for GMI.

1. Introduction

1.1. Background and related work

Gun model identification (GMI) is one of today's essential research topics. GMI has a crucial role in criminalistics, and it has been mostly used in military applications. At the same time, GMI can also be used for security purposes in applications such as digital forensics and forensics [1–3]. Each gun has a special acoustic characteristic. When these acoustic characteristics are analyzed in detail, they can provide critical support information to criminalistics. Different features such as the audios obtained from the trigger and hammer mechanism of the gun, the audio of mechanical movement, the audios of the bullet hitting solid surfaces customize the gun [4,5]. Therefore, the detection of this gun can be achieved by using the features of a gun audio signal. GMI systems with a high recognition rate are needed in the crime scene to recognize such systems automatically [6]. These systems must be capable of responding to events in different environments. In addition, systems should be less sensitive to environmental sounds. For example, a gunshot may be environmentally inadequate in acoustical evidence due to its location [7]. Therefore, the audio sample should be recorded in high quality and evaluated. Sampling gunfire from different sources is also useful in obtaining evidence [8–10]. Obtaining evidence from a gunshot is a complex process because these conditions are not always performed. Criminalistics always try to obtain evidence in a military incident. For this purpose, automatic recognition systems have been developed to be used in different applications such as GMI, gun caliber, and position [11,12].

Artificial intelligence is applied to different disciplines today [13–19]. The purpose of the application of artificial intelligence is to reach meaningful information, especially in the rapidly increasing multimedia data. The primary purpose of using artificial intelligence techniques in this study is to determine the gun model

E-mail address: sdogan@firat.edu.tr

https://doi.org/10.1016/j.apacoust.2021.107916
0003-682X/

used in an incident at a crime scene. Different studies are carried out for this purpose in the literature. Some of these studies are given as follows.

Sallai et al. [20] presented a method for weapon classification. Data were collected for evaluation in this study. The data collected consists of 4 days and 194 gunshots. The accuracy rate was computed as 95.00% in the study. Morton et al. [21] demonstrated a gunshot signatures detection model using Hidden Markov Model. 5 guns and 46 gunshots were used in the study. The classification rate was computed as 95.65%. Djeddou and Touhami [22] proposed a feature selection method for weapon classification. This study was based on the Gaussian mixture model. In this study, 14 guns and 230 gunshots were collected. The classification rate was presented as 96.29%. Bajzik et al. [23] developed a gunshot detection algorithm. This study was based on convolutional neural networks. In this study, the performance of the system with noisy gunshots was evaluated. TensorFlow and Keras frameworks have been evaluated for this purpose, and the accuracy rate was calculated as 99.14% with ResNet18. Rahman et al. [24] demonstrated an automatic gunshots detection method. In this study, data were collected from YouTube [25]. This data consists of 778 samples. For different classifiers, the accuracy rate was calculated as 94.97%. Raponi et al. [11] presented a method for gunshot classification. In this study, the convolutional neural network and short-time Fourier transform were used and data collected from YouTube. The data consists of 59 guns and 3655 gunshots. In this study, accuracy rate was obtained as 90.00%. Sanchez-Hevia et al. [26] proposed an approach for gunshot detection. The main purpose of this work is to assist the multichannel acoustic detection of gun. In this study, GUNS - Construction Kit [27] dataset was used. This dataset consists of 14 guns and 840 gunshots. According to the evaluation results, the accuracy rate was 94.10%. Souli and Lachiri [28] proposed a new method for security applications. The proposed method can be used especially for audio surveillance applications. In the proposed method, a study has been performed to examine the focused audio in situations involving environmental sounds. The proposed method is based on principal component analysis and scattering transform. Stowell et al. [29] presented an approach to detailed analysis of audio scenes involving environmental sounds using machine learning techniques. A dataset was created to test the performance of the approach developed in this study. This dataset has been published for different studies. In this study, accuracy rate was calculated as 77.00%. Papadimitriou [30] et al showed a model for audio-based event detection. This model provides a detailed analysis of the raw audio. Mel-Frequency Cepstral Coefficients method was preferred for this analysis. At the same time, Short-Time Fourier Transform method was used to increase the recognition accuracy.

1.2. Motivation and our method

Audio forensics is one of the fundamental branches of digital forensics. The audio forensics processes are generally implemented by using an expert, and the experts collect evidence from audios manually. Therefore, audio forensics processes are time-consuming processes. In order to overcome this problem, artificial intelligence or machine learning-based audio forensics tools should be developed.

This research focuses on solving GMI using gunshot audios. Audio-based GMI is cheaper than ballistic and chemometrics methods. Also, ballistic and chemometrics methods are time-consuming processes. An appropriate and high accurate audio classification method can identify the gun model in a few seconds. Thus, an automated GMI method is presented, and a new corpus is collected. This corpus has 2310 gunshots audios of the 28 gun models.

A high accurate classification method must have an effective feature generation method to extract discriminative patterns. Simple and effective methods must be used to present a high accurate, and lightweight learning model. Here, hand-crafted feature generation functions are used. The hand-crafted features are generally divided into grouped as textural and statistical. The presented method uses both textural and statistical features. A fractal pattern is used to present the textural feature generation function. TQWT [31] is an effective decomposition method for one-dimensional signals and can detect differences easily. Therefore, it is utilized as a decomposer, and levels are created deploying TQWT. Textural and statistical feature generators extract low-level features. However, they can extract high-level features using a multileveled feature generation structure. NCA [32] selects the most meaningful features, and these features are classified using kNN [33] or SVM [34,35]. These classifiers are shallow classifiers. The objective of using shallow classifiers is to denote the discriminative attributes of the generated and chosen features.

1.3. Novelties and contributions

The novelties of the presented model are;

* A new gunshot audio dataset is collected as a testbed.
* New fractal feature generation model is proposed in this paper. The recommended feature generation function is used by H-tree fractal. Therefore, it is named as H-tree fractal pattern.

Contributions;

* Variable patterns have been presented to generate discriminative features in the literature. This work uses a fractal tree for feature generation. The obtained results showed that the presented fractal pattern-based model attained high performance. This work clearly implies that the fractal geometric shapes can be used for feature generation, and more models can be presented. Moreover, a multileveled feature generation network is created using textural (textural features are generated using a fractal H-tree pattern) and statistical features. By applying these, a fused/hybrid feature generator is presented, and this feature generation model can be deployed to solve other signal classification problems.
* GMI is one of the most complex issues for forensics science. In order to overcome this problem, an automated gunshot classification model must be presented. Therefore, a gunshot audio dataset and an accurate gunshot sound classification model are presented.

2. Gunshot audio dataset

Novel gunshot audio dataset is collected from the YouTube video-sharing platforms. This corpus contains 2310 audios of the 28 gun models. The primary objective of this corpus is to present a new testbed for audio forensics (http://web.firat.edu.tr/sdogan/Gun_S_Dogan.rar). The collected audios are sampled with 44.1 kHz and 48 kHz. They have two channels. However, the first channel was used in this work. The file extension of them is mp3. The collected dataset is a heterogeneous dataset [25]. The attributes of this dataset are given in Table 1.

Table 1
The attributes of the collected gunshot audio dataset.

Id	Model	Instances	Duration (second)
1	AK47	73	152
2	Baby desert	91	186.01
3	Baretta 92 fs	41	73.13
4	Colt 1911	43	70.90
5	CZ-75-SP-01	53	39.11
6	Desert eagle	100	200
7	EEA-Witness	27	86.33
8	Famas	84	164.70
9	FN Brownin Hi-Power	79	138.10
10	FN Five Seven	107	234.97
11	FNX-Nine	79	215.97
12	Glock G17	76	150.84
13	Imi Tavor 21	69	158.23
14	AK12	98	1147
15	M16	100	200
16	M249	99	198
17	MG42	100	200
18	MKE-MPT-55	44	108.77
19	MP5	100	200
20	MPT76	57	100.57
21	PAP MP92	82	92
22	Sa-80-a2	39	97.96
23	Sig sauer P226	100	163.66
24	Springfield Armory	81	185.46
25	Taurus PT 92	64	43.88
26	The-FN-Scar	92	156.33
27	USP Compact	62	218
28	Walther P38	90	155.97
Total		2130	5137.89

3. The presented sleep stage classification model

3.1. Overview of the presented model

This paper recommends a fractal pattern based gunshot audio classification method to identify gun models. Multileveled TQWT and fused feature generation, selection of the most informative features, and the selected features are classified using kNN or SVM classifiers. The general steps of this method are given below.

Step 0: Read audio data.

Step 1: Extract statistical features deploying the used 18 statistical moments.

Step 2: Extract textural features using the presented fractal H-tree pattern. This function generates 320 features

Step 3: Concatenate the generated features.

Step 4: Apply TQWT to the used audio signal and obtain sub-bands. TQWT has three parameters, and these parameters are named q-factor (Q), redundancy value (r), level number (L). Here, 1,2, and 8 are selected as parameters of the TQWT. TQWT generates L + 1 sub-bands(*SB*). Therefore, *nine* sub-bands are obtained in this step.

Step 5: Apply Steps 1–3 to the generated each sub-band by TQWT.

Step 6: Merge the generated features from each level and obtain $3380 = (320 + 18) * 10$ features from audio.

Step 7: Deploy NCA to the merged features.

Step 8: Select the most informative 683 features from the generated 3280 features.

The schematic explanation of the presented model is shown in Fig. 1.

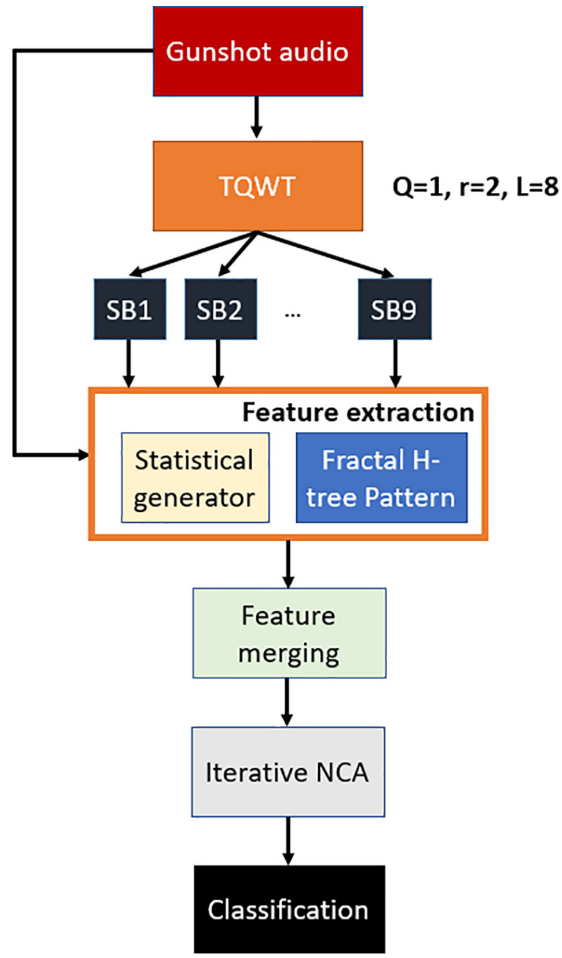

Fig. 1. Schematic explanation of the presented fractal H-tree based GMI model using gunshot audios.

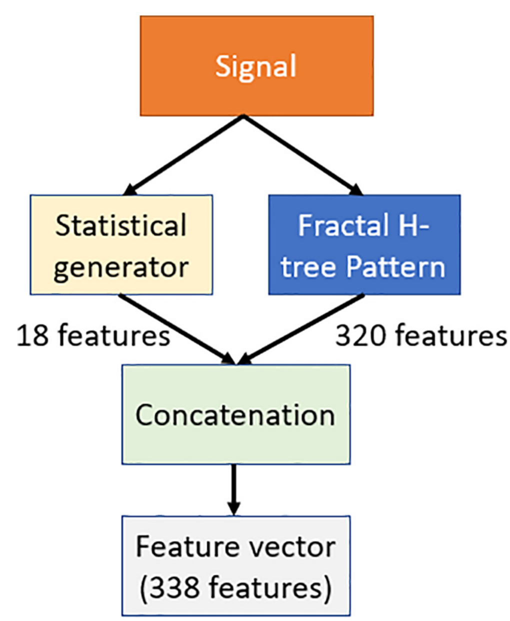

Fig. 2. Snapshot of the presented fused feature generation.

3.2. Feature generation

The feature generation is one most essential phases of the presented method. Here, both textural and statistical feature extraction functions are used together. A multileveled feature generation structure is used to generate high-level features. By applying the TQWT decomposition method, nine sub-bands are obtained, and the presented textual and statistical feature

Table 2
The used 18 statistical moments to generate statistical features.

1	Mean
2	Median
3	Standard deviation (std)
4	Root mean square error
5	Skewness
6	Kurtosis
7	Minimum
8	Maximum
9	Range (Minimum-maximum)
10	Absolute maximum (max(abs))
11	Absolute minimum (min(abs))
12	Absolute mean (mean(abs))
13	Absolute median(median(abs))
14	Absolute standard deviation
15	Energy
16	Shannon entropy
17	Sure entropy
18	Log Entropy

generation functions are applied to these sub-bands and raw audio signal. In each level/section, 338 features are generated. 320 of the 338 features are textural features, and 18 of them are statistical features. The graphical demonstration of the presented fused feature generation is shown in Fig. 2.

The presented feature generation method is defined below as mathematically.

$$X(j) = merge(St(Signal), Htree - Pat(Signal)), j = \{1, 2, \cdots, 338\} \quad (1)$$

$$X(k*338+j) = merge\left(St\left(SB^k\right), Htree - Pat(Signal)\right), k = \{1, 2, \cdots, 9\} \quad (2)$$

where $merge(.,.)$ defines merging function, $St(.)$ is a statistical feature generator, and $Htree - Pat(.)$ is the recommended fractal H-tree pattern-based textural feature generator. $St(.)$ and $Htree - Pat(.)$ extract 18 and 320 features consecutively, X is generated feature vector with a length of 3380. The used functions are explained below.

3.2.1. Statistical feature generation

The statistical feature generation functions have been widely preferred for the hand-crafted methods, and variable statistical moments have been used. In this work, linear and nonlinear statistical generators are used. The used statistical moments are listed in Table 2 [36].

3.2.2. The presented fractal H-tree pattern

A new fractal pattern based function is presented to generate textural features. This function uses fractal H-tree and signum function together. The fractal theory is one of the hot-topic research areas. Repeated shapes are created using fractals, and complex geometric shapes are created. A new fractal shape is created using the H tree, and the created H tree-based fractals are shown in Fig. 3.

The presented fractal H-tree pattern uses n = 2 fractal. A 7 × 7 sized matrix is used to model this fractal as a pattern. The used nodes (red circles) define nodes, and black lines are represented as edges. The edges show the relation between the values (nodes). The used fractal H-tree pattern is shown in Fig. 4.

The steps of the recommended fractal H-tree pattern are;

0: Read signal.

1: Divide signal into 49 sized overlapping blocks.

$$b^i(j) = Signal(i + j - 1), i = \{1, 2, \cdots, L - 48\}, j = \{1, 2, \cdots, 49\} \quad (3)$$

where b^i is i[th] block, *Signal* is the used one-dimensional signal with a length of L.

2: Deploy vector to matrix transformation and obtain 7 × 7 sized vector to implement the recommended fractal H-tree pattern.

$$m^i(h, t) = b^i(j), h = \{1, 2, \cdots, 7\}, t = \{1, 2, \cdots, 7\} \quad (4)$$

where m^i defines the i[th] matrix with a size of 7 × 7.

3: Assign the used values per the used fractal H-tree pattern (See Fig. 4).

4: Generate bits using the presented pattern and signum function. Mathematical descriptions of the binary feature generation are given below.

$$\begin{vmatrix} bit^1(1) \\ bit^1(2) \\ bit^1(3) \\ bit^1(4) \\ bit^1(5) \\ bit^1(6) \end{vmatrix} = Sgn\left(\begin{vmatrix} m(1,1), m(2,2) \\ m(2,1), m(2,2) \\ m(3,1), m(2,2) \\ m(1,3), m(2,2) \\ m(2,3), m(2,2) \\ m(3,3), m(2,2) \end{vmatrix}\right) \quad (5)$$

$$\begin{vmatrix} bit^2(1) \\ bit^2(2) \\ bit^2(3) \\ bit^2(4) \\ bit^2(5) \\ bit^2(6) \end{vmatrix} = Sgn\left(\begin{vmatrix} m(1,5), m(2,6) \\ m(2,5), m(2,6) \\ m(3,5), m(2,6) \\ m(1,7), m(2,6) \\ m(2,7), m(2,6) \\ m(3,7), m(2,6) \end{vmatrix}\right) \quad (6)$$

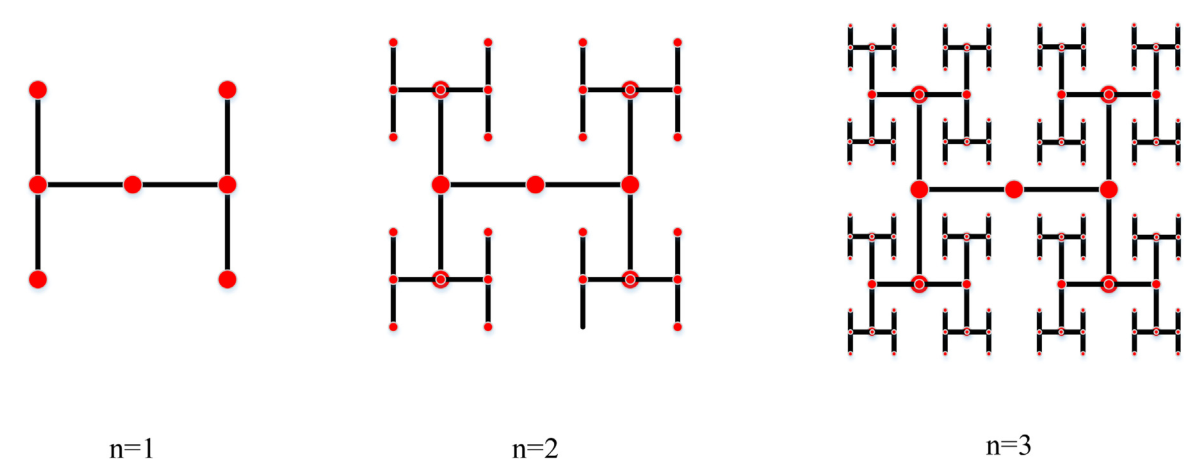

Fig. 3. The presented H tree fractals.

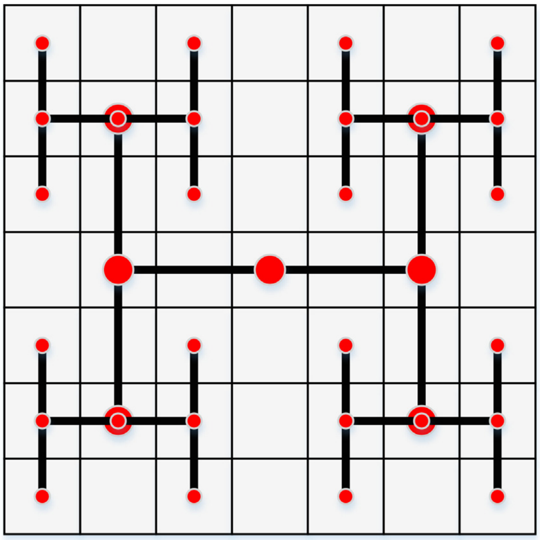

Fig. 4. The used fractal H-tree pattern. Here, there are five H trees. Each H tree has a center value. By deploying these five center values, the other six values, and the signum function. 320 features are extracted.

$$\begin{vmatrix} bit^3(1) \\ bit^3(2) \\ bit^3(3) \\ bit^3(4) \\ bit^3(5) \\ bit^3(6) \end{vmatrix} = Sgn\left(\begin{vmatrix} m(5,1), m(6,2) \\ m(6,1), m(6,2) \\ m(7,1), m(6,2) \\ m(5,3), m(6,2) \\ m(6,3), m(6,2) \\ m(7,3), m(6,2) \end{vmatrix}\right) (7)$$

$$\begin{vmatrix} bit^4(1) \\ bit^4(2) \\ bit^4(3) \\ bit^4(4) \\ bit^4(5) \\ bit^4(6) \end{vmatrix} = Sgn\left(\begin{vmatrix} m(5,5), m(6,6) \\ m(6,5), m(6,6) \\ m(7,5), m(6,6) \\ m(5,7), m(6,6) \\ m(6,7), m(6,6) \\ m(7,7), m(6,6) \end{vmatrix}\right) (8)$$

$$\begin{vmatrix} bit^5(1) \\ bit^5(2) \\ bit^5(3) \\ bit^5(4) \\ bit^5(5) \\ bit^5(6) \end{vmatrix} = Sgn\left(\begin{vmatrix} m(2,2), m(4,4) \\ m(4,2), m(4,4) \\ m(6,2), m(4,4) \\ m(2,6), m(4,4) \\ m(4,6), m(4,4) \\ m(6,6), m(4,4) \end{vmatrix}\right) (9)$$

$Sgn(fp, sp) = \begin{cases} 0, fp - sp \geq 0 \\ 1, fp - sp < 0 \end{cases}$ (10)where $Sgn(.,.)$ describes signum function, fp represents the first parameter of the signum function, and sp is the second input parameter of the signum function, and m is the used matrix.

Step 4: Calculate five map signals using the generated bits.

$map^k(i) = \sum_{y=1}^{6} bit^k(y) * 2^{y-1}, k = \{1, 2, \cdots, 5\}$ (11)

where map^k defines the kth map signal.

5: Create five histograms, and each histogram has 64 values because the generated map signals are coded by 6-bits.

$hist^k(map^k(i)) = hist^k(map^k(i)) + 1$ (12)where $hist^k$ is the kth histogram with a size of 64.

6: Merge the generated histogram to obtain the feature vector ($fvec$) with a length of 320.

$fvec((k-1) * 64 + l) = hist^k(l), l = \{1, 2, \cdots, 64\}$ (13)

The given six steps are defined as our presented fractal H-tree pattern ($Htree - Pat(.)$).

3.3. Feature selection

In the feature selection phase, the NCA selector is used. NCA is a distance and weight-based selector. NCA calculates weights using a Manhattan distance-based weight calculation function. It generates non-negative weights for each feature. However, the selection of the optimal features by deploying NCA is a hard process. To solve this problem, Tuncer et al. [37] presented an iterative variation of the NCA, and it is called INCA. This research generates 3380 features. All of the features are utilized as the input of the NCA. Then, the weights of each feature are calculated by NCA. B using the calculated weights, the iterative feature selection process is implemented. An error value generator should be used in the iterative feature selection. kNN classifier is utilized as an error/loss value generator. Steps of this phase are;

1: Calculate the weight of each feature and find indices (idx) of the sorted weights to select the most valuable features

2: Select the feature vectors using the generated weights iteratively.

$fv^j(d, j) = X(d, idx(j)), d = \{1, 2, \cdots, noI\}, j = \{1, 2, \cdots, 3380\}$ (14)

where fv^j defines the jth feature vector, and noI defines a number of instances.

3: Find the error value of each selected feature vector using a kNN classifier with 10-fold cross-validation.

$loss^j = kNN(fv^j, target, 1, MD, 10)$ (15)where $loss^j$ is jth loss value, $kNN(.,.,.,.,.)$ defines kNN classifier and parameter of it are features, target, k value, distance metric, and k value of fold-validation respectively.

4: Calculate the index of the minimum loss value ($inde$).

5: Select the optimal feature (fv^{opt}) deploying $inde$ and idx.

$fv^{opt}(d, j) = X(d, idx(j)), j = \{1, 2, \cdots, inde\}$ (16)

By applying this method, 683 features are selected as fv^{opt}.

3.4. Classification

Classification is the last phase of the presented fractal H-tree pattern-based GMI method. Here, two shallow classifiers are used, and they are kNN and SVM. The selected 683 features are forwarded to these classifiers to predict the model of the guns using gunshot audios. Parameters of the used classifiers are listed in Table 3.

4. Experiments

In the evaluation of the recommended fractal H-tree based method, the MATLAB programming environment is used. The recommended method was coded using functions. We defined the

Table 3
The parameters of the used classifiers.

Parameters	kNN	SVM
Kernel	–	3rd-degree polynomial
Validation	10-fold	10-fold
k value	1	–
Box constraint level	–	1
Multiclass method	–	One-vs-One
Voting	None	–
Distance	Manhattan	–

Table 4
The calculated UAR, AP, F1, and GM results per classifier.

Classifier	UAR	AP	F1	GM
kNN	96.02%	96.48%	96.20%	95.96%
SVM	90.20%	91.84%	90.84%	89.99%

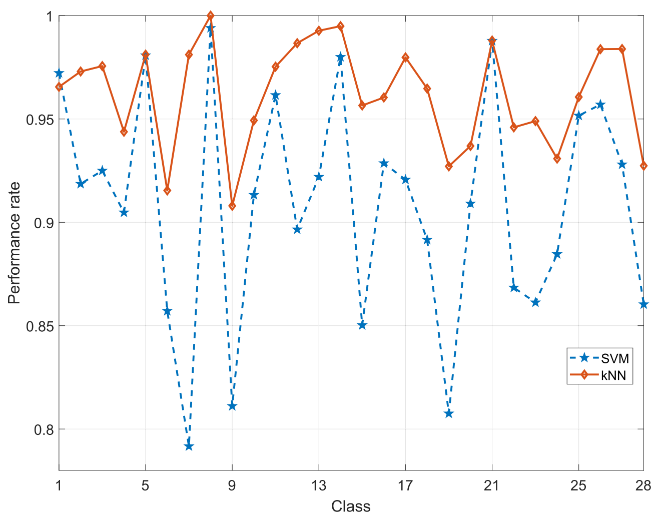

Fig. 5. The calculated F1-scores class by class using kNN and SVM classifiers.

main H-tree, statistical feature generation, Iterative NCA selector, and classification functions. The main function is called the others to implement our method. The recommended approach is applied to the collected audios, and results are calculated using 10-fold cross-validation. Unweighted average recall (UAR), average precision (AP), F1-score (F1), and geometric mean (GM) were used to present results clearly [38].

Table 4 denotes the presented model yielded high-performance metrics for both classifiers. Accuracy is one of the widely used performance evaluation metrics in the literature. Here, 10-fold cross-validation is utilized as a training and testing method. Therefore, accuracy results

Also, the calculated F1-scores (class by class) are shown in Fig. 5.

5. Discussions

Gunshot audio classification is a complex issue for machine learning and digital forensics. An intelligent audio forensics tool should be developed to help audio forensics experts. This research focuses on the GMI problem. A new gunshot audio dataset was collected from the open-source videos to overcome this problem. This dataset has 2310 audios of the 28 gun models. The recommended method generates both textural and statistical features multileveled. By deploying this method, both low and high levels features are extracted. The used INCA selected the most valuable features. The feature selection process of this research is shown in Fig. 6.

As stated in Fig. 6, the length of the optimal feature vector is 683. These 683 features are forwarded to kNN and SVM classifiers. 96.10% and 90.85% accuracy are yielded. This model is compared to other state-of-art models to show the success of this model on the GMI. The calculated accuracies are listed in Table 6.

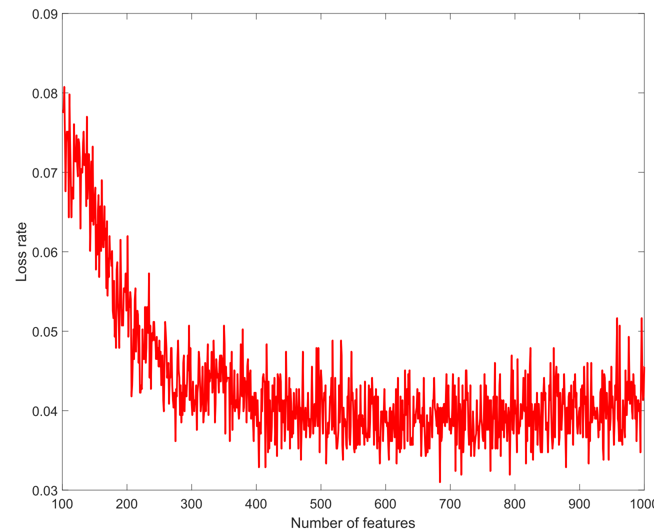

Fig. 6. NCA based feature selection process in the range of [100,1000].

Table 5
Fold by fold accuracies (%) of the used classifiers.

Fold	kNN	SVM
Fold-1	96.71	94.84
Fold-2	98.59	91.55
Fold-3	99.53	93.90
Fold-4	96.24	86.38
Fold-5	92.02	86.38
Fold-6	96.24	93.90
Fold-7	93.43	88.26
Fold-8	96.24	91.55
Fold-9	97.65	91.55
Fold-10	94.37	90.14
Overall	96.10	90.85

Table 6
The comparison results.

Studies	Year	Dataset	Accuracy (%)
Sanchez-Hevia et al. method [39]	2015	Collected data (14 weapon-1680 individuals registers)	56.90
Kiktova et al. method [40]	2015	JDAE TUKE [41] (4 guns-372 gunshots)	80.00
Sanchez-Hevia et al. method [26]	2017	GUNS - Construction Kit [27] (14 guns-840 gunshots)	94.10
Roma et al. method [42]	2018	D-CASE2013 [43] (320 samples)	60.00
Souli and Lachiri method [28]	2018	GTZAN dataset [44] (224 samples)	89.23
Ozkan and Barkana method [1]	2019	DASE database [45] (332 samples)	60.19
Rahman et al. method [24]	2020	Collected data (778 samples of gunshot and 778 normal scream)	94.97
Papadimitriou et al. method [30]	2020	MIVIA Audio Events Dataset (8112 gunshot extend) [46]	95.21
Raponi et al. method [11]	2020	Collected data (59guns-3655 gunshots)	90.00
The proposed method		Collected data (28 guns-2130 gunshot)	96.10

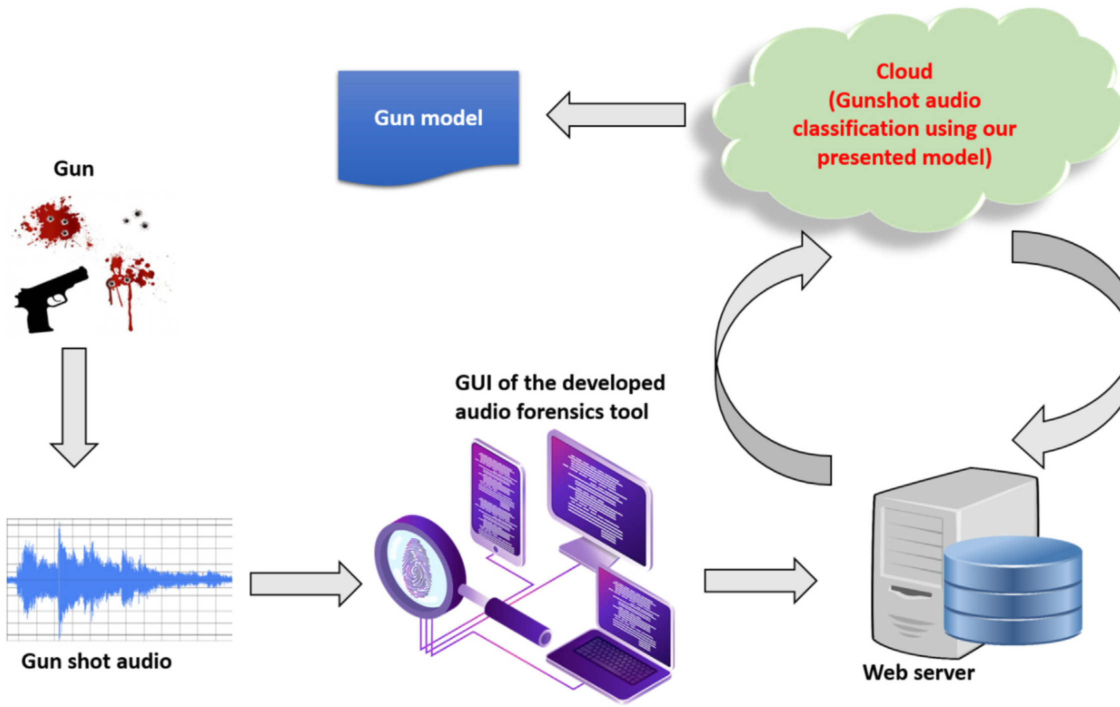

Fig. 7. The planned automated GMI tool using our presented gunshot audio classification method. Here, gunshot audios are collected from the crime scene, and these audios are given to the developed audio forensics tool. This tool is a web-based tool, and a trained dataset is loaded to the cloud. By using this trained big dataset, gunshot audios are tested, and the calculated results (gun models) are reported.

Table 5 denotes the success of the presented fractal H-tree pattern and NCA based gunshot classification method for GMI. Also, this model was tested on a big dataset. As stated in Table 6, the best result was calculated as 96.10%. Advantages of this research are;

* The recommended fractal H-tree pattern is a new fractal textural feature extraction function. The presented model reached high classification accuracies on a big dataset. It is obviously demonstrated the success of the fractal geometry-based feature generation function.
* GMI is a problem for digital forensics. Machine learning-based methods should be used to solve this problem. However, there are limited corpora in the literature to train and test the machine learning methods. Therefore, a new big dataset was acquired.
* By deploying basic methods/algorithms, an effective and high accurate gunshot audio classification model is presented. Therefore, this method can program using a computer with simple configurations.
* Our fractal feature generation based method reached greater performance than other methods using a big dataset.
* Robust results yielded using 10-fold cross-validation. By applying this method, an automated GMI method can be presented for real-world detection.

I am planning to present a new project to develop a new digital forensics toolbox for GMI using gunshot audios deploying our presented methods. Also, new fractal-based textural feature extraction functions can be presented because there are variable fractals in nature. The primary objective of this method is to present a nonparametric learning method for one-dimensional signals. The presented fractal H-tree-based method can be employed to classify other one-dimensional signals, and an improved deep version can be used instead of recurrent neural networks. The snapshot of my intended project is shown in Fig. 7.

6. Conclusion

This research recommends both a high accurate automated GMI method and a big dataset (gunshot audios dataset). The presented corpus contains 2310 gunshot audios of the 28 gun models. The recommended method uses a fractal-based feature generation function. A multileveled feature generation method is presented by deploying this fractal-based function (fractal H-tree pattern), statistical features, and TQWT. The recommended feature generation method extract features in both the frequency domain and time domain. NCA selects the most informative features. These features are classified using kNN and SVM. The presented method yielded 96.10% and 90.85% classification accuracies by employing kNN and SVM, respectively. Per the comparison table, the recommended method yielded high success on a bigger dataset than others.

CRediT authorship contribution statement

Sengul Dogan: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is supported by Firat University Research Fund, Turkey Project Number: TEKF.20.09.

References

[1] Ozkan Y, Barkana BD. Forensic audio analysis and event recognition for smart surveillance systems. In: 2019 IEEE International Symposium on Technologies for Homeland Security (HST): IEEE; 2019. p. 1–36.

[2] Chen C-Y, Abdallah A, Wolf W. Audiovisual gunshot event recognition. In: 2006 IEEE International Conference on Systems, Man and Cybernetics: IEEE; 2006. p. 4807–12.

[3] Paraskevas I, Rangoussi M. Feature extraction for audio classification of gunshots using the Hartley transform; 2012.

[4] Ahmed T, Uppal M, Muhammad A. Improving efficiency and reliability of gunshot detection systems. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing: IEEE; 2013. p. 513–7.

[5] Freire IL, Apolinário Jr JA. Gunshot detection in noisy environments. In: Proceeding of the 7th International Telecommunications Symposium, Manaus, Brazil2010. p. 1–4.

[6] Busse C, Krause T, Ostermann J, Bitzer J. Improved Gunshot Classification by Using Artificial Data. Audio Engineering Society; 2019.

[7] Pikrakis A, Giannakopoulos T, Theodoridis S. Gunshot detection in audio streams from movies by means of dynamic programming and bayesian networks. 2008 IEEE International Conference on Acoustics, Speech and Signal Processing: IEEE 2008:21–4.

[8] Uzkent B, Barkana BD, Cevikalp H. Non-speech environmental sound classification using SVMs with a new set of features. Int J Innov Comput Inf Control 2012;8:3511–24.

[9] Maher RC. Acoustical characterization of gunshots. In: 2007 IEEE Workshop on Signal Processing Applications for Public Security and Forensics: IEEE; 2007. p. 1–5.

[10] Valenzise G, Gerosa L, Tagliasacchi M, Antonacci F, Sarti A. Scream and gunshot detection and localization for audio-surveillance systems. In: 2007 IEEE Conference on Advanced Video and Signal Based Surveillance: IEEE. p. 21–6.

[11] Raponi S, Ali I, Oligeri G. Sound of Guns: Digital Forensics of Gun Audio Samples meets Artificial Intelligence. arXiv preprint arXiv:200407948; 2020.

[12] Showen R, Calhoun R, Chu WC, Dunham J. Acoustic gunshot location in complex environments: concepts and results. Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense VII: International Society for Optics and Photonics; 2008. p. 694318.

[13] Hashimoto DA, Witkowski E, Gao L, Meireles O, Rosman G. Artificial intelligence in anesthesiology current techniques. Clin Appl Limitations. Anesthesiol: J Am Soc Anesthesiol 2020;132:379–94.

[14] Díez-Sanmartín C, Sarasa CA. Application of artificial intelligence techniques to predict survival in kidney transplantation: a review. J Clin Med 2020;9:572.

[15] Jahani A, Rayegani B. Forest landscape visual quality evaluation using artificial intelligence techniques as a decision support system. Stoch Env Res Risk Assess 2020;34:1473–86.

[16] Shariati M, Mafipour MS, Mehrabi P, Shariati A, Toghroli A, Trung NT, et al. A novel approach to predict shear strength of tilted angle connectors using artificial intelligence techniques. Eng Comput 2020;1–21.

[17] Sisodiya N, Dube N, Thakkar P. Next-generation artificial intelligence techniques for satellite data processing. In: Artificial Intelligence Techniques for Satellite Image Analysis: Springer; 2020. p. 235–54.

[18] Aydemir E, Gulsecen S. Arranging bus behaviour by finding the best prediction model with artificial neural networks. Tehnički vjesnik 2019;26:885–92.

[19] Aydemir E. Weka ile yapay zeka. Ankara: Seçkin Yayınevi; 2018.

[20] Sallai J, Hedgecock W, Volgyesi P, Nadas A, Balogh G, Ledeczi A. Weapon classification and shooter localization using distributed multichannel acoustic sensors. J Syst Archit 2011;57:869–85.

[21] Morton Jr KD, Torrione PA, Collins L. Classification of acoustic gunshot signatures using a nonparametric Bayesian signal model. In: Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense X: International Society for Optics and Photonics; 2011. p. 80190T.

[22] Djeddou M, Touhami T. Classification and modeling of acoustic gunshot signatures. Arab J Sci Eng 2013;38:3399–406.

[23] Bajzik J, Prinosil J, Koniar D. Gunshot detection using convolutional neural networks. In: 2020 24th International Conference Electronics: IEEE; 2020. p. 1–5.

[24] Rahman SU, Khan A, Abbas S, Alam F, Rashid N. Hybrid system for automatic detection of gunshots in indoor environment. Multimedia Tools Appl 2020;1–11.

[25] Youtube. https://www.youtube.com/; 2020.

[26] 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 Trans Audio Speech Lang Process 2017;25:1172–82.

[27] BOOM. Library GbR, "GUNS—Construction Kit," Available: http://www.boomlibrary.com/boomlibrary/products/guns; 2011.

[28] Souli S, Lachiri Z. Audio sounds classification using scattering features and support vectors machines for medical surveillance. Appl Acoust 2018;130:270–82.

[29] Stowell D, Giannoulis D, Benetos E, Lagrange M, Plumbley MD. Detection and classification of acoustic scenes and events. IEEE Trans Multimedia 2015;17:1733–46.

[30] Papadimitriou I, Vafeiadis A, Lalas A, Votis K, Tzovaras D. Audio-based event detection at different SNR settings using two-dimensional spectrogram magnitude representations. Electronics 2020;9:1593.

[31] Selesnick IW. Wavelet transform with tunable Q-factor. IEEE Trans Signal Process 2011;59:3560–75.

[32] Raghu S, Sriraam N. Classification of focal and non-focal EEG signals using neighborhood component analysis and machine learning algorithms. Expert Syst Appl 2018;113:18–32.

[33] Maillo J, Ramírez S, Triguero I, Herrera F. kNN-IS: An Iterative Spark-based design of the k-Nearest Neighbors classifier for big data. Knowl-Based Syst 2017;117:3–15.

[34] Vapnik V. The support vector method of function estimation. Nonlinear Modeling: Springer; 1998. p. 55–85.

[35] Vapnik V. The nature of statistical learning theory. Springer science & business media; 2013.

[36] Kuncan F, Kaya Y, Kuncan M. New approaches based on local binary patterns for gender identification from sensor signals. J Faculty Eng Arch Gazi Univ 2019;34.

[37] Tuncer T, Dogan S, Özyurt F, Belhaouari SB, Bensmail H. Novel multi center and threshold ternary pattern based method for disease detection method using voice. IEEE Access 2020;8:84532–40.

[38] Tuncer T, Aydemir E. An automated local binary pattern ship identification method by using sound. Acta Infologica. 2020;4:57–63.

[39] Sánchez-Hevia HA, Ayllón D, Gil-Pita R, Rosa-Zurera M. Gunshot classification from single-channel audio recordings using a divide and conquer approach. In: ICPRAM (2); 2015. p. 233–40.

[40] Kiktova E, Lojka M, Pleva M, Juhar J, Cizmar A. In: Gun type recognition from gunshot audio recordings. 2015): IEEE; 2015. p. 1–6.

[41] Pleva M, Vozáriková E, Dobos Lu, Cizmár A. The joint database of audio events and backgrounds for monitoring of urban areas. J Electr Electron Eng 2011;4:185.

[42] Roma G, Herrera P, Nogueira W. Environmental sound recognition using short-time feature aggregation. J Intell Inf Syst 2018;51:457–75.

[43] Giannoulis D, Benetos E, Stowell D, Rossignol M, Lagrange M, Plumbley MD. Detection and classification of acoustic scenes and events: An IEEE AASP challenge. In: 2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics: IEEE; 2013. p. 1–4.

[44] Tzanetakis G, Cook P. Musical genre classification of audio signals. IEEE Trans Speech Audio Process 2002;10:293–302.

[45] Barkana BD, John N, Saricicek I. Auditory suspicious event databases: DASE and Bi-DASE. IEEE Access 2018;6:33977–85.

[46] Foggia P, Petkov N, Saggese A, Strisciuglio N, Vento M. Reliable detection of audio events in highly noisy environments. Pattern Recogn Lett 2015;65:22–8.