Turkish Journal of Science & Technology
20(1), 63-76, 2025

Research Paper
https://doi.org/10.55525/tjst.1571845

A Novel Car Interior Sound Classification Method based on Multileveled Local Binary Four Patterns and Iterative ReliefF

Erhan AKBAL[1*], Sengul DOGAN[2], Turker TUNCER[3]
[1,2,3] Department of Digital Forensics Engineering, Technology Faculty, Firat University, Elazig, Turkey

[*1] erhanakbal@firat.edu.tr, [2] sdogan@firat.edu.tr, [3] turkertuncer@firat.edu.tr

(Geliş/Received: 22/10/2024; Kabul/Accepted: 23/12/2024)

Abstract: Sound classification is one of the crucial study areas in machine learning and sound forensics. However, there are limited studies on sound forensics or sound-based crime investigations in the digital forensics literature. In this work, a novel area of sound forensics is presented: car interior sound classification (CISC). The main aim of CISC is to identify a car using its interior environmental sound. A car interior sound dataset was collected using 10 car models. This CISC model includes feature generation using the local binary four pattern and one-dimensional multilevel discrete wavelet transform (DWT), iterative ReliefF-based feature selection, and classification. k-nearest neighbors (kNN) and support vector machine (SVM) were utilized as classifiers to demonstrate the general success of the proposed learning model for CISC. The accuracy rates were calculated as 93.72% ± 0.37 and 95.04% ± 0.30 with kNN and SVM, respectively. These results demonstrate the success of the proposed method.

Key words: Car interior sound classification, Iterative ReliefF, Local Binary Four Pattern, Environmental sound classification, Sound forensics Analysis, Cyber Crime.

Çok Seviyeli Yerel İkili Dört Desenler ve İteratif ReliefF Tabanlı Yeni Bir Araç İçi Ses Sınıflandırma Yöntemi

Öz: Ses sınıflandırması, makine öğrenimi ve ses adli bilişiminde önemli çalışma alanlarından biridir. Ancak, dijital adli bilişim literatüründe ses adli bilişimi veya ses tabanlı suç soruşturmaları üzerine sınırlı sayıda çalışma bulunmaktadır. Bu çalışmada, ses adli bilişiminde yeni bir alan sunulmaktadır: araç içi ses sınıflandırması (CISC). CISC'nin temel amacı, araçların iç ortam seslerini kullanarak tanımlanmasıdır. Bu amaçla, 10 farklı araç modeli kullanılarak bir araç içi ses veri seti oluşturulmuştur. CISC modeli, yerel ikili dört desen ve tek boyutlu çok seviyeli ayrık dalgacık dönüşümü (DWT) ile özellik çıkarımını, iteratif ReliefF tabanlı özellik seçimini ve sınıflandırmayı içermektedir. Modelin genel başarısını göstermek için k-en yakın komşu (kNN) ve destek vektör makinesi (SVM) sınıflandırıcıları kullanılmıştır. kNN ve SVM ile elde edilen doğruluk oranları sırasıyla %93,72 ± 0,37 ve %95,04 ± 0,30 olarak hesaplanmıştır. Bu sonuçlar, önerilen yöntemin başarısını ortaya koymaktadır.

Anahtar kelimeler: Araba iç ses sınıflandırması, döngüsel rölyef, yerel ikili dört desen, çevresel ses sınıflandırması, ses adli bilişimi analizi, siber suç.

1. Introduction

In recent years, many environmental sound classification (ESC) methods have been presented for signal processing and digital forensics. There are many studies in the literature for determining environmental sounds. Studies generally cover topics such as environment monitoring [1, 2], health applications [3], diagnostic systems [4-8], environment recognition [9], cyber security [10, 11], fault detection [12, 13] and different ambient sounds [14, 15]. Generally, common studies are in the field of ESC [16-20]. In digital forensics, many sound data are obtained from the digital storage device [10, 21, 22]. The sounds are listened by the examiner for a detailed understanding of the sounds obtained. The frequency values that the human ear can hear are limited. Human auditory system cannot detect some evidence in sound format [23]. While law enforcement officers follow criminals, they usually use ways such as listening to the environment and listening to the phone. Obtaining information such as the location and environment of the criminal from a sound data can provide great convenience in terms of following the suspect. For instance, when there is no information about the suspect, law enforcement officers have difficulties following up [24]. Follow-up can be made easier with the predict of the crime scene.

---

[*] Corresponding author: erhanakbal@firat.edu.tr. ORCID Number of authors: [1] 0000-0002-5257-7560, [2] 0000-0001-9677-5684, [3] 0000-0002-5126-6445

Automatic sound classification methods aim to achieve fast results by minimizing human errors. Thus, at the moment of the incident, fast and error-free information about the crime is obtained [25, 26].

There are many classification studies in the literature using local binary pattern. Some of them are given as follows. Güner et al. [27] showed that LBP can be used to classification of different modulation types. Gupta et al. [28] used an optimized binary pattern algorithm to classify white blood cells. Hu et al. [29] demonstrated the performance of LBP for the classification of pump vibrations for IoT systems.

Classification of environmental sounds is a subject of signal processing [30, 31]. Many different methods have been used in signal processing, such as deep learning, machine learning, and artificial neural networks [32-37]. The aim of all studies is to present high accurate sound classification method for sound signals [38]. There are different studies in the literature on sound classification/detection using artificial intelligence methods and some of these are; Fan et al. [39] proposed a model for classification different environmental sounds for use in hearing aids. It is aimed to automatically classification 5 different environmental sounds (bus, subway, street, indoor and car) using deep neural networks. They proposed algorithm has achieved a classification rate of 98.8%. Jaber et al. [40] suggested a classification algorithm for disease detection using lung sounds. Random Forest, AdaBoost, Gradient Boosting algorithms are used. The dataset consists of 5 classes and 99.04% accuracy classification rate is shown. Shen et al. [41] proposed a system for classification of sound events obtained from real environments. The proposed system uses Gaussian mixture models and Mel-Frequency Cepstral Coefficients (MFCCs). Accuracy rate was calculated as 91.36% for 8 classes. Saki and Kehtarnavaz [42] proposed a method for real-time classification of sound signals obtained from hearing aids. The three-level model is intended to distinguish music, speech and noise. With their model, it was demonstrated that traditional classification models could be reached high accuracies. Abdoli et al. [16] presented a classification model based on 1D Convolution Neural Network (CNN). Their model resulted 89% accuracy on UrbanSound8k dataset [43]. Medhat et al. [44] applied the deep CNN method of the spectrograms of the used sound signals. Masked ConditionaL Neural Network (MCLNN) model was applied on ESC-10 and ESC-50 datasets [22]. Chen et al. [45] emphasized that the CNN model used in ESC caused information loss during pooling. In their study, a model to eliminate data loss during pooling was presented. Souli et al. [46] presented an approach to the classification of environmental sounds. Scatter transform and principal component analysis (PCA) were used to extract the feature vectors of the sound signals. It achieved 92.22% classification performance using SVM classifier. López-Pacheco et al. [47] suggested a method for classification of urban sounds. Orthogonal Matching Pursuit (OMP) algorithm was used as a feature generation method. The results obtained showed that it was successful in determining the dominant sounds in urban dataset. Tuncer et al. [48] presented a model for disease detection in sound signals. Their model consisted of feature extraction with 1D-LBPNet, informative feature selection with NCA and classification using 1NN. Sounds were classified with an accuracy of 98.83%. AlQahtani [49] presented a model to facilitate forensic investigations. Experimental studies showing the problems of sound recognition environments about the model were presented. The K-Nearest Neighbors classifier was utilized as classifier. Muhammad and Alghathbar [10] presented an experimental study on sound media detection in digital forensics. MFCCs algorithm was employed as feature generator. In addition, MPEG-7 audio dataset was proposed. Their dataset contained 10 classes and they reached to 75% from 90% classification accuracies.

1.1. Motivation of this work

In this paper, our main aims are to define a sub-branch of ESC and develop a novel sound forensics method. Therefore, we used car interior sounds, and a novel CISC method is presented. We collected a sound dataset, which contains 700 sounds with 10 classes. To accurately classify this dataset, a novel learning model is presented in this work.

1.2. Our Method

In this study, a novel CISC method is proposed using sounds obtained from the 10 most commonly used C-class cars in the world market. This dataset contains 10 different car models, and each class has 70 sound signals, totaling 700 sounds. Our method is multileveled and includes multileveled feature generation with LBFP, iterative feature selection with ReliefF, and classification phases. Our fundamental goals are outlined as follows. We used a 7-level one-dimensional discrete wavelet transform to generate low, medium, and high-level features. LBFP uses four variable patterns to comprehensively generate features, with 544 features extracted at each level. A total of 544 x 8 = 4352 features are generated (LBFP is applied to the raw sound signal and 7 low-pass filters of it). The most valuable features are selected using IRF, and the selected features are then forwarded to classifiers.

1.3. Contributions

Our contributions are given as below.

- In this paper, we define a novel ESC-like study area, which can be considered a sub-branch of ESC. The study area we present is car interior sound classification (CISC). CISC is directly dependent on the automotive industry, digital forensics, signal processing, and machine learning. To propose a novel model for CISC, we collected a car interior sounds (CIS) dataset.
- A novel multileveled feature generation method is presented using a new feature generation approach. The proposed LBFP uses four variable patterns to comprehensively extract features from CIS. ReliefF faces an optimal feature selection problem; therefore, iterative ReliefF is introduced to choose the optimal number of features. SVM and kNN were utilized as classifiers. The proposed LBFP and IRF-based method achieved high success rates using conventional classifiers. This demonstrates the discriminative strength of the LBFP and IRF-based feature extraction and selection processes. A classification accuracy of 95.86% was achieved using this method for CISC.

2. Materials

We curated CIS from YouTube point-of-view (PoV) drive videos [50]. Road conditions are given as follows. The sounds of these cars were collected on dry and asphalt roads. We selected 10 widely preferred car models, and the acoustics of these cars were recorded using the Windows audio recorder in m4a format. These sounds were segmented into 3–5-second frames, and their frequency is 48 KHz. The segmentation of the sounds into 3–5-second frames was deliberately chosen to balance the need to capture meaningful audio information. Variations in segment length accommodate differences in the acoustic events captured in each recording. For example, longer segments can include transitions or continuous sounds necessary for context, while shorter segments ensure that isolated or discrete sounds are not lost in the analysis. The interval also aligns with the duration of steady-state in-car sounds under consistent conditions (e.g., driving on a paved road). This approach minimizes the risk of introducing noise or losing important details due to excessively short or excessively long segments. By defining a flexible yet controlled segmentation interval, the dataset is designed to maintain high quality for feature extraction and classification of the in-car sound environment. Additionally, this approach helps create a balanced dataset. In some cars, the collected interior sounds are longer than in others. Therefore, sound segments of varying lengths were used.

The attributes of the collected CIS dataset are listed in Table 1. These cars use petrol (gasoline) engines and manual transmissions.

Table 1. Attributes of the collected CIS dataset.

No	Model	Observation	No	Model	Observation
1	Clio 4 - 2018	70	6	Meganne 4 - 2018	70
2	Fiat Egea 1.4 - 2018	70	7	Passat B8 - 2018	70
3	Ford Focus Mk4- 2018	70	8	Peugeot 308 - 2019	70
4	Golf VII - 2019	70	9	Skoda Octavia - 2019	70
5	Honda Civic - 2018	70	10	Toyota Corolla XI - 2018	70

As seen in Table 1, the collected CIS dataset has 700 sounds. This dataset named as CISC10. The graphical demonstration of the samples of the CISC10 is shown in Figure 1.

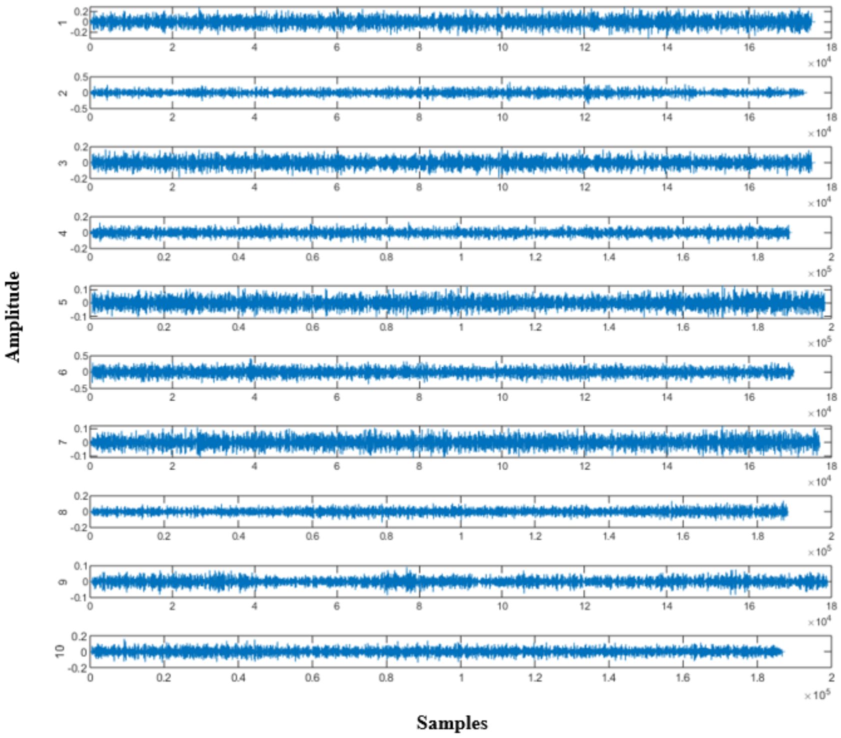


Figure 1. Graphical demonstration of the samples of CISC10 dataset according to classes.

3. The proposed CISC model

A novel learning model is presented in this paper. Our model uses the proposed LBFP for feature generation, IRF to select meaningful features and SVM [51], kNN [52, 53] for classification. The used feature generation phase is multileveled. In order to achieve, low, medium and high features should be generated. Therefore, we used one dimensional multilevel DWT [54] for decomposition. Multileveled DWT is applied on CIS and seven low pass coefficients are obtained. LBFP extracts 544 features from each low pass coefficients and raw CIS. These features are concatenated and 4352 features are generated. IRF selects 693 most valuable from the generated 4352 features. These 693 features are utilized as input of the kNN and SVM classifiers and predictive results are calculated. The proposed LBFP and IRCA based CISC method is summarized as Figure 2.

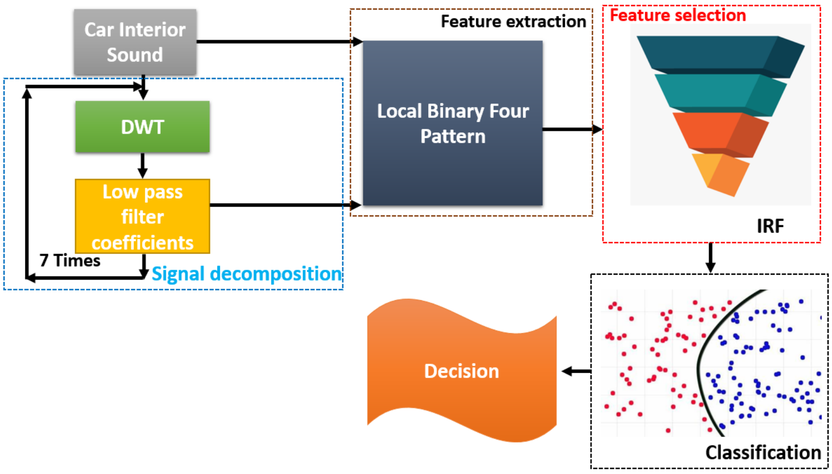


Figure 2. Demonstration of the block diagram of our proposed CISC method.

The used phases and steps of the proposed method are explained in subsections.

3.1. The proposed LBFP based multileveled feature generation

Our first phase is LBFP based multileveled feature generation. In this phase, multilevel DWT decomposition [55, 56], feature extraction with LBFP and feature concatenation processes are used. Details of LBFP also in this section.

Step 0: Load CIS.
Step 1: Apply multilevel 1D-DWT on the loaded CIS. As it can be seen from Equations 1-2.

$$[low_1\ high_1] = dwt(CIS, sym4) \tag{1}$$

$$[low_{k+1}\ high_{k+1}] = dwt(low_k, sym4), k = \{1,2, \ldots, 6\} \tag{2}$$

where low_k and $high_k$ denote kth leveled low pass filter and high pass filter coefficients, $sym4$ represnets symlets 4 filter. This filter has been mostly preferred signal processing application for decomposition and noise reduction. Therefore, $sym4$ filter is chosen and this filter is one of the commonly used noise reduction wavelet filters. The selection of level 7 in DWT decomposition is aimed at obtaining the most optimal level of features and frequency content of in-car audio signals in accordance with the sound dimension. Higher decomposition levels effectively capture low-frequency, steady-state sounds specific to vehicle interiors by separating frequency components in greater detail. At level 7, the approximation coefficients retain sufficient detail to represent these low-frequency components while discarding irrelevant high-frequency noise. Although detail coefficients are widely used in signal classification and feature engineering, in this study, the approximation coefficients are utilized to focus on the salient acoustic features critical for in-car audio classification. This selection also minimizes noise originating from the nature of the signal. Furthermore, experiments have demonstrated that the level 7 approximation coefficients combined with the proposed LBFP method provide the highest classification performance among various parameter configurations.

Step 2: Extract 544 features from CIS and low pass filter coefficients by using LBFP.

$$feat^1 = LBFP(CIS) \tag{3}$$

$$feat^t = LBFP(low_{t-1}), t = \{2,3, \ldots, 8\} \tag{4}$$

$feat^t$ is tth feature vector.

As it can be seen Equations 3-4, the fundamental function of the feature generation is LBFP. LBFP is a LBP like feature extractor. It uses four variable patterns and these patterns extract 256, 256, 32 and 32 features respectively. Finally, these features are fussed and 544 features are obtained. LBFP procedure is explained in below steps.

Step 2.1: Divide CIS into 9 sized overlapping windows. As it can be seen from Equation 5.

$$window^h = CIS(i{:}\, i + 8), i = \{1,2, \ldots, L\}, h = \{1,2, \ldots, L - 8\} \tag{5}$$

where $window^h$ hth 9 sized overlapping windows, L represents length of the CIS.

Step 2.2: Generate binary features using signum function and the defined 4 patterns. The used patterns are shown in Figure 3.

These patterns (See Figure 2) are used to generate features with signum function. Equation of signum function is shown in Equation 6.

$$sgnm(P_N, P_C) = \begin{cases} 0, P_N - P_C < 0 \\ 1, Otherwise \end{cases} \tag{6}$$

where $sgnm(.,.)$ is signum function, P_N and P_C denote first and second parameters of the signum function respectively. Explanation of the bit generation procedure of the proposed LBFP is shown in Equations 7-10.

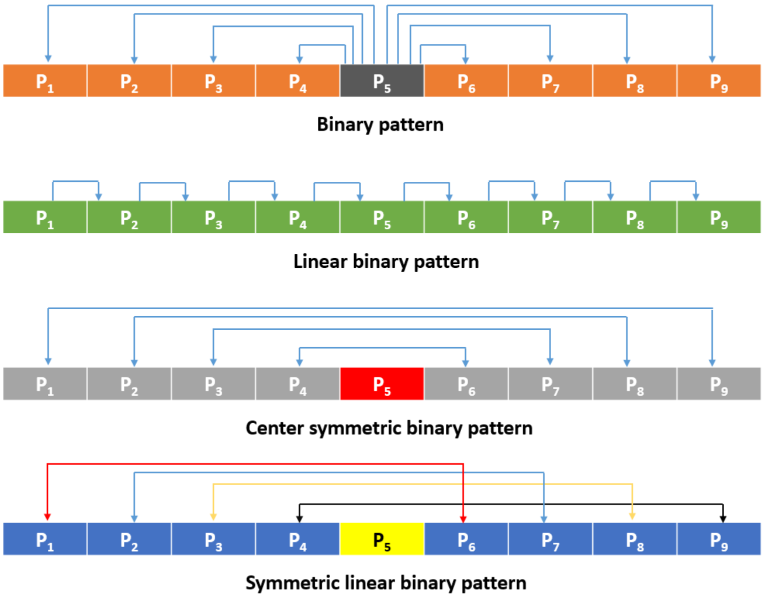


Figure 3. The used patterns in the proposed LBFP.

$$bit^1(i) = sgnm\big(window(i), window(5)\big), i = \{1,2,\dots,9\}, i \notin \{5\} \quad (7)$$

$$bit^2(j) = sgnm\big(window(j), window(j+1)\big), i = \{1,2,\dots,8\} \quad (8)$$

$$bit^3(g) = sgnm\big(window(g), window(10-g)\big), g = \{1,2,3,4\} \quad (9)$$

$$bit^4(g) = sgnm\big(window(g), window(g+5)\big) \quad (10)$$

where bit^1, bit^2, bit^3 and bit^4 denote extracted bits with binary pattern (BP) [57], linear binary pattern (LiBP), center symmetric binary pattern (CSBP) and symmetric linear binary pattern (SLBP) respectively. i, j and g values represent index values. As it can be seen from Equations 7-10, BP, LiBP, CSBP and SLBP extracts 8, 8, 4 and 4 bits respectively.

Step 2.3: Construct four map values using the extracted bits and binary to decimal value conversion. As it can be seen from Equations 11-14.

$$mv^1(h) = \sum_{i=1}^{8} bit^1(i) * 2^{8-i} \quad (11)$$

$$mv^2(h) = \sum_{i=1}^{8} bit^2(i) * 2^{8-i} \quad (12)$$

$$mv^3(h) = \sum_{i=1}^{4} bit^3(i) * 2^{4-i} \quad (13)$$

$$mv^4(h) = \sum_{i=1}^{4} bit^4(i) * 2^{4-i} \quad (14)$$

where mv^k is k^th^ map values.

Step 2.4: Generate histograms of the map values. In this step, four histograms are calculated. The length of these histograms are calculated as $2^8, 2^8, 2^4$ and 2^4 for BP, LiBP, CSBP and SLBP values respectively. Therefore, four arrays with size of 256, 256, 16 and 16 are defined and initial values of them are assigned as 0. Histograms are generated by using Equations 15-17.

$$histo^t\left(mv^t(h)\right) = histo^t\left(mv^t(h)\right) + 1, t = \{1,2,\dots,4\} \quad (15)$$

where $histo^t$ defines histogram of t^th^ map value.

Step 2.5: Concatenate the generated four histograms to obtain feature vector ($featvec$) with length of 544.

$$featvec = histo^1 \cup histo^2 \cup histo^3 \cup histo^4 \quad (16)$$

where ∪ denotes concatenation operator.

Step 3: Concatenate feature vector.

$$X = feat^1 \cup feat^1 \cup \dots \cup feat^8 \quad (17)$$

where X is concatenated feature vector with size of 4352.

Step 1-3 have been defined the recommended multilevel feature extraction method. This feature extraction method is designed to capture the distinct features of in-car sounds. The MDWT decomposes audio signals into multiple levels, enabling the extraction of frequency components while reducing the noise ratio at each level. As a result, features are obtained in both spatial and frequency domains. Additionally, DWT serves as both a preprocessing and a multilevel feature extraction function, creating a multilevel feature extraction method similar to deep learning approaches. In this phase, LBFP is employed as the main feature extractor.

The presented LBFP method is inspired by local binary patterns (LBP) and is specifically designed to extract robust and distinctive features. By utilizing four unique patterns—binary pattern (BP), linear binary pattern (LiBP), center-symmetric binary pattern (CSBP), and symmetric linear binary pattern (SLBP)—LBFP captures hidden patterns in overlapping signal windows. Each pattern provides a different perspective on the signal structure, enabling the extraction of nuanced, texture-like features from a 1D audio signal. For instance, BP captures local differences between sample points, while CSBP and SLBP extract more global and symmetric relationships.

The proposed method leverages the complementary strengths of both approaches by combining DWT with LBFP. DWT performs multi-resolution analysis, filtering out noise and focusing on meaningful frequency components, while LBFP extracts descriptive features from the processed signals. Experimental results demonstrate that integrating DWT and LBFP in this manner achieves optimal performance.

3.2. Iterative ReliefF feature selector

Relief [58] has been mostly used feature selector in the literature and improved version of the Relief is named as ReliefF. Both Relief and ReliefF [59] generates weights. Relief uses Euclidean distance-based fitness function but ReliefF uses Manhattan distance-based fitness function to calculate optimal weights of each features. In this work, ReliefF is selected as feature selector. The generated negative weighted features can be assigned as redundant feature. However, number of optimal features has been selected parametrically. To automatically select number of optimal features, iterative ReliefF (IRF) is chosen as feature selector. To calculate error values, we need a classifier. kNN is selected as error value calculator in this work. Steps of our IRF based feature selection are given as below.

Step 4: Apply IRF function to generated features (X) for selecting final features use Equation 18.

$$feat^S = IRF(X, y) \quad (18)$$

where $feat^S$ represents selected features, $IRF(.,.)$ defines the used supervised feature selection function and y denotes target.
Steps of the IRF function are;

Step 4.1: Apply ReliefF function to features (X) for calculation weights and sorted indices of the weights. Equation 19 showed below.

$$[weight\ index] = RF(X, y) \quad (19)$$

where RF is ReliefF function, $weight$ ReliefF weights and $index$ is sorted indices.

Step 4.2: Set lower and upper bounds for number of features. This step is applied to decrease time cost of the IRF. In this work, lower and upper bounds are selected as from 100 to 1100 respectively.

Step 4.3: Calculate error values iteratively used to Equations 20,21.

$$feature^{t-99}(i) = X\big(index(i)\big), t = \{100,101, \dots ,1100\}, i = \{1,2, \dots , t\} \tag{20}$$

$$error(t-99) = kNN(feature^{t-99}, y, 1, 10, MD) \tag{21}$$

The parameters of the kNN functions are given as follows. These are features, target, k value, cross validation parameter (10-fold CV is used), and distance metric (MD is manhattan distance).

Step 4.4: Calculate minimum error value and index of it with Equation 22.

$$[minimum, index^{min}] = min\ (error) \tag{22}$$

Step 4.5: Select optimal number of features by using index of minimum error ($index^{min}$) used Equation 23.

$$feat^{S}(i) = X\big(index(i)\big), i = \{1,2, \dots , index^{min} + 99\} \tag{23}$$

In this work, IRF selected 693 features. The plotting of the calculated error values with IRF is shown in Figure 4.

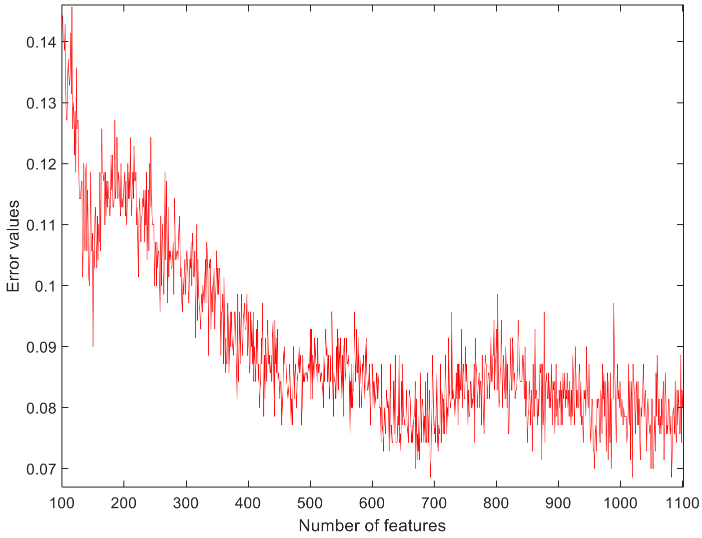


Figure 4. Error values of the IRF by using kNN classifier.

3.3. Classification

The classification phase is the final phase of the proposed method. The numerical performance metrics are calculated using classifiers. In this section, kNN and SVM classifiers are employed. The attributes of these classifiers are as follows: the kNN classifier used is the 1NN classifier, which uses the Manhattan distance for classification. The other classifier is SVM, which has various mathematical kernels. In this work, we used a second-degree polynomial kernel to obtain classification results. The constraint level (C) is set to 2, and the one-vs-all method is chosen for multiclass classification. 10-fold cross-validation (CV) is one of the most preferred training and testing strategies in the literature, and therefore, it is used to obtain the results.

Step 5: Classify selected 693 features with kNN or SVM classifiers with 10-fold CV.

4. Experimental results

Results were shown in this section. We used two classifiers to calculate numerical results. In the classification papers, F1-score ($F1$), geometric mean ($gmean$) and classification accuracy (CA) have been widely used to obtain numerical results of the used classifiers. To calculate these values, number of true positives (ntp), true negatives (ntn), false positives (nfp) and false negatives (nfn) values should be used. We implemented the proposed multileveled LBFP and IRF on a desktop computer by using MATLAB2018a programming environment. By using MATLAB Classification Learner Tool (MCLT), kNN and SVM were executed and we generated code of the used classifiers and $F1$, $gmean$ and CA codes were added these classifiers codes. Explanation of these evaluation criteria were given as Equations 24-26 [60].

$$F1 = \frac{2ntp}{2ntp + nfn + nfp} \tag{24}$$

$$gmean = \sqrt{\frac{ntp * ntn}{(ntp + nfn)(ntn * nfp)}} \tag{25}$$

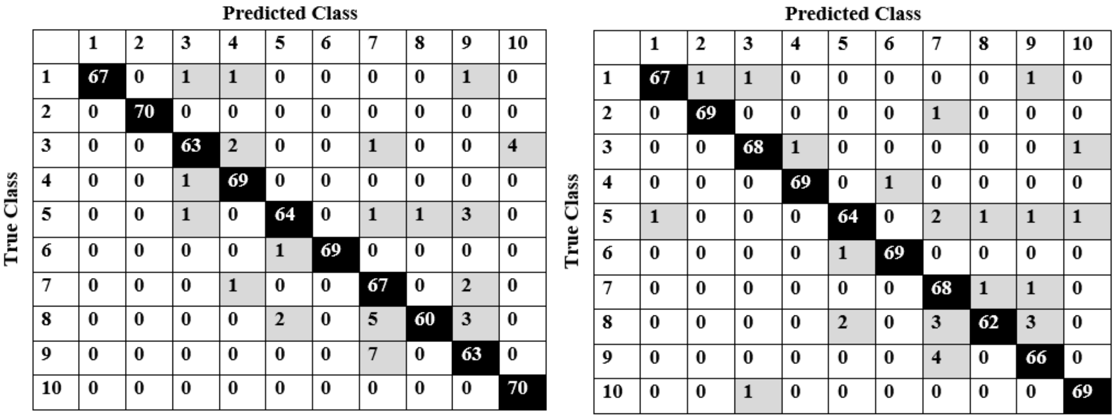
$$CA = \frac{ntp + ntn}{ntp + ntn + nfp + nfs} \tag{26}$$

The calculated performance criteria for each classifier were listed in Table 2. To calculate general success rates (average value ± standard deviation of the value), test process of each classifier was executed 1000 times.

Table 2. F1-score (%), geometric mean (%) and accuracy scores of the used classifiers.

Classifier	Evaluation Type	F1-score	Geometric mean	Accuracy
kNN	General	93.90±0.36	93.57±0.39	93.72±0.37
	Maximum	94.75	94.47	94.57
SVM	General	95.10±0.29	94.96.10±0.30	95.04±0.30
	Maximum	95.92	95.80	95.86

Confusion matrices of the best results were shown in Figure 5.



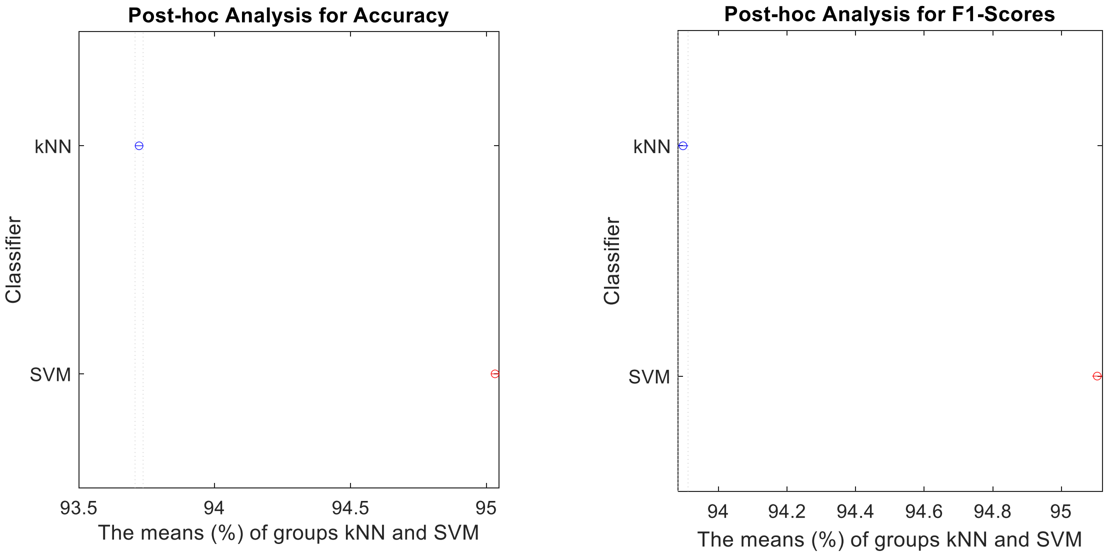
	1	2	3	4	5	6	7	8	9	10
1	67	0	1	1	0	0	0	0	1	0
2	0	70	0	0	0	0	0	0	0	0
3	0	0	63	2	0	0	1	0	0	4
4	0	0	1	69	0	0	0	0	0	0
5	0	0	1	0	64	0	1	1	3	0
6	0	0	0	0	1	69	0	0	0	0
7	0	0	0	1	0	0	67	0	2	0
8	0	0	0	0	2	0	5	60	3	0
9	0	0	0	0	0	0	7	0	63	0
10	0	0	0	0	0	0	0	0	0	70

	1	2	3	4	5	6	7	8	9	10
1	67	1	1	0	0	0	0	0	1	0
2	0	69	0	0	0	0	1	0	0	0
3	0	0	68	1	0	0	0	0	0	1
4	0	0	0	69	0	1	0	0	0	0
5	1	0	0	0	64	0	2	1	1	1
6	0	0	0	0	1	69	0	0	0	0
7	0	0	0	0	0	0	68	1	1	0
8	0	0	0	0	2	0	3	62	3	0
9	0	0	0	0	0	0	4	0	66	0
10	0	0	1	0	0	0	0	0	0	69

Figure 5. a) Confusion matrix of the best result of the kNN b) Confusion matrix of the best result of the SVM.

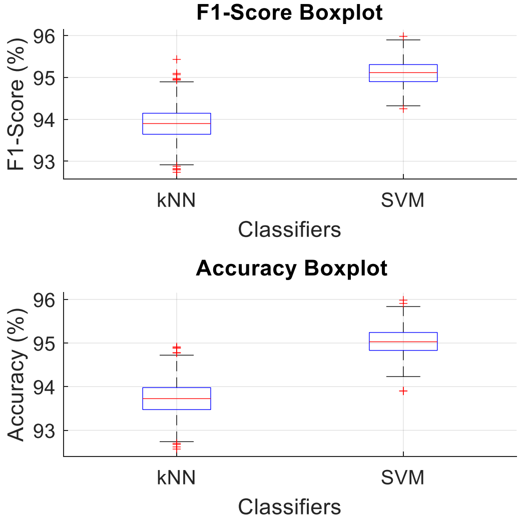
5. Discussions and conclusions

In this work, we presented a novel research area called CISC, which can be defined as a sub-branch of ESC. ESC is crucial for digital forensics, signal processing, and machine learning. The main aim of CISC is to identify a car using its cabin sound. To achieve this, we collected a novel CIS dataset. This dataset contains 700 sounds from 10 widely preferred car models. To demonstrate the feasibility of CISC, we introduced novel methods: the LBFP feature extractor and the IRF feature selector. Multilevel DWT and LBFP were used together to generate low, medium, and high-level features. The most discriminative ones were selected using IRF, and the selected features were forwarded to kNN and SVM classifiers. Classification accuracies of 95.86% and 94.57% were achieved using SVM and kNN classifiers, respectively. The confusion matrices of the best results were also presented. As shown in these matrices, while the best result for kNN is lower than that of SVM, kNN achieved 100.0% classification accuracy for two classes. Other widely used performance criteria were also employed. The general success rates (%) of kNN were calculated as 93.90±0.36, 93.57±0.39, and 93.72±0.37 for F1-score, geometric mean, and accuracy, respectively. For SVM, the values were 95.10±0.29, 94.96±0.30, and 95.04±0.30 for F1-score, geometric mean, and accuracy, respectively. The statistical analysis of the results has been given Figure 6.



(a) Mean accuracy

(b) Mean F1-score



(c) Accuracy and F1-scores of the presented models

Figure 6. The statistical analysis of the computed classification accuracies and F1-scores.

The p-value for both F1-score and accuracy in the ANOVA test is 0.0003 (approximated to a very small value due to numerical precision limits). This indicates that the differences in the performance metrics between the kNN and SVM classifiers are significant.

To clearly demonstrate the success of the proposed LBFP and IRF-based sound classification method, comparative results are listed in Table 3.

Table 3. Comparison of the proposed classification method with other sound classification methods.

Method	Aim and Methods	Performance	Number of class
[27]	Automatic digital modulation classification using LBP	95%	6 classes
[39]	ESC for implementation on hearing aid app using deep neural network	98.8%	5 classes
[29]	Classification of pumps vibration and sounds using varying classifier and feature extractors	99.47%	3 classes
[40]	Lung sounds classification using ensemble classifier algorithms	99.04%	7 classes
[45]	ESC using dilated convolutions	78%	10 classes
[16]	ESC using 1D CNN	89%	10 classes
[46]	Audio sound classification for medical surveillance using SVM	92.22%	10 classes
[10]	Environment recognition for audio forensics using MFCCs	96%	4 classes
[6]	Acoustic-based fault diagnosis of roller bearings using a deep graph convolutional network	92%	4 classes
[12]	Termite detection system based on acoustic using various ML algorithms	93.83%	2 classes
[13]	Fault diagnosis of bearing and stator faults using acoustic signals using Nearest Neighbor, Nearest Mean classifier and GMM	95.3%	3 classes
Our Method	Car interior sound classification by using LBFP and IRF	95.92%	10 classes

As seen in Table 3, previously presented 10-class classification methods achieved classification accuracies of 78%, 89%, and 92.22%, respectively. In contrast, our method achieved a 95.92% success rate on a new dataset with 10 classes.

The key points of the recommended model are:

- The SVM classifier achieved higher general and maximum performance metrics compared to kNN. Specifically, SVM's maximum accuracy reached 95.86%, compared to 94.57% for kNN.
- Both classifiers show strong diagonal dominance in the confusion matrices (Figure 5), indicating accurate predictions for most classes.
- SVM showcases fewer misclassifications overall compared to kNN, particularly for classes with more complex patterns, such as classes 8 and 9.
- Certain classes, such as Class 10, achieved perfect classification accuracy in kNN (70/70 correct classifications). Similarly, SVM also yielded high precision for Class 10 but with a single misclassification.
- For challenging classes like Class 8, SVM showed better performance with fewer misclassifications (3 errors compared to 5 in kNN).
- The low standard deviations across multiple runs (±0.36 for kNN and ±0.29 for SVM in F1-score) indicate that both classifiers are robust and provide consistent results across multiple test iterations.
- The multilevel feature extraction and iterative feature selection contributed significantly to high classification performance. The integration of approximation coefficients from DWT and descriptive features from LBFP provides meaningful feature representation.
- While SVM performs better in terms of overall metrics and generalization, kNN demonstrated strong performance for certain classes, especially those with less variation in features, such as Classes 1, 6, and 10.
- For higher accuracy and robustness, SVM is preferable, whereas kNN may be sufficient for less computationally intensive tasks.
- The hierarchical extraction of features through DWT and LBFP significantly boosts the discriminative power of the classifiers.

- Misclassifications often occur between similar classes (e.g., Classes 5 and 6 or Classes 7 and 8), likely due to overlapping or less distinct features in these cases. Future work could focus on enhancing inter-class separability by incorporating additional feature extraction techniques or using ensemble models.

These results clearly demonstrate the general success of the proposed method and the feasibility of CISC. The benefits and novelties of the proposed LBFP and IRF-based CISC method are listed below.

- A novel sound classification area was defined.
- The feasibility of CISC was demonstrated using a novel learning model.
- A new one-dimensional feature extractor, called LBFP, was introduced.
- The optimal feature selection problem of ReliefF was solved using IRF.
- A highly accurate CISC method was proposed, and its success was demonstrated using two classifiers.

6. Future directions

Our proposals for future works are;

- People spend a significant amount of time in cars. Therefore, the car interior is crucial for digital forensics and crime investigation, as it can also be considered a crime scene. Consequently, many methods and datasets can be developed for CISC.
- LBFP is a one-dimensional feature extractor. By using LBFP, other signal-related problems can also be addressed.
- Other branches of sound forensics or ESC can be defined.
- Novel sound forensics tools and applications can be implemented using the proposed LBFP and INCA-based methods for crime investigators.
- In the near future, autonomous cars and vehicles will likely become common in traffic. These cars will be intelligent, and to authenticate them, novel intelligent CISC methods can be developed.

References

[1] Lou SC, Chao RM, Ko SH, Lin KM and Zhong JX. A simplified signal analysis algorithm for the development of a low cost underwater echo-sounder. Measurement 2011; 44(9):1572-1581.

[2] Liu S. Measurement and analysis of vibration and noise in the ambient environment of metro. Measurement 2020; 107998.

[3] Amarnath M, Sugumaran V and Kumar H. Exploiting sound signals for fault diagnosis of bearings using decision tree. Measurement 2013; 46(3):1250-1256.

[4] Lara R, Jimenez-Romero R, Perez-Hidalgo F and Redel-Macias M D. Influence of constructive parameters and power signals on sound quality and airborne noise radiated by inverter-fed induction motors. Measurement 2015; 73:503-514.

[5] Ye S, Zhang J, Xu B, Song W, and Zhu S. Experimental studies of the vibro-acoustic characteristics of an axial piston pump under run-up and steady-state operating conditions. Measurement 2019; 133: 522-531.

[6] Zhang D, Stewart E, Entezami M, Roberts C and Yu D. Intelligent acoustic-based fault diagnosis of roller bearings using a deep graph convolutional network. Measurement 2020; 156:107585.

[7] Xiang M, Zang J, Wang J, Wang H, Zhou C, Bi R and Xue C. Research of heart sound classification using two-dimensional features. Biomed Signal Proces 2023; 79:104190.

[8] Chen J, Guo Z, Xu X, Zhang L B, Teng Y, Chen Y and Wang W. A robust deep learning framework based on spectrograms for heart sound classification. IEEE/ACM Trans Comput Biol Bioinform 2023; 21(4): 936-947

[9] Petošić A, Djurek I, Suhanek M and Grubeša S. Interlaboratory comparisons' measurement uncertainty in the field of environmental noise. Measurement 2019; 148: 106932.

[10] Muhammad G and Alghathbar K. Environment recognition for digital audio forensics using MPEG-7 and mel cepstral features. J Electr Eng 2011; 62(4):199-205.

[11] Ratcliffe JH, Lattanzio M, Kikuchi G, Thomas K. A partially randomized field experiment on the effect of an acoustic gunshot detection system on police incident reports. J Exp Criminol 2019; 15(1):67-76.

[12] Nanda MA, Seminar KB, Nandika D, Maddu A. Development of termite detection system based on acoustic and temperature signals. Measurement 2019; 147: 106902.

[13] Glowacz A, Glowacz W, Glowacz Z, Kozik J. Early fault diagnosis of bearing and stator faults of the single-phase induction motor using acoustic signals. Measurement 2018; 113:1-9.

[14] Kane P and Andhare A. Critical evaluation and comparison of psychoacoustics, acoustics and vibration features for gear fault correlation and classification. Measurement 2020;107495.

[15] Mohaimenuzzaman M, Bergmeir C, West I T, Meyer B. Environmental Sound Classification on the Edge: A Pipeline for Deep Acoustic Networks on Extremely Resource-Constrained Devices. Pattern Recognit 2023;133:109025.

[16] Abdoli S, Cardinal P, Koerich AL. End-to-end environmental sound classification using a 1d convolutional neural network. Expert Syst Appl 2019; 136:252-263.

[17] Ahmad S, Agrawal S, Joshi S, Taran S, Bajaj V, Demir F, Sengur A. Environmental sound classification using optimum allocation sampling based empirical mode decomposition. Phys A Stat 2020; 537: 122613.

[18] Bardou D, Zhang K, Ahmad SM. Lung sounds classification using convolutional neural networks. Artif Intell Med 2018; 88: 58-69.

[19] Beltrán J, Chávez E, Favela J. Scalable identification of mixed environmental sounds, recorded from heterogeneous sources. Pattern Recognit Lett 2015;68:153-160.

[20] İnik Ö, CNN hyper-parameter optimization for environmental sound classification. Appl Acoust 2023; 202:109168.

[21] Khunarsal P, Lursinsap C, Raicharoen T. Very short time environmental sound classification based on spectrogram pattern matching. Inf Sci 2013; 243:57-74.

[22] Piczak, KJ. ESC: Dataset for environmental sound classification. In: Proceedings of the 23rd ACM international conference on Multimedia; 2015; New York, NY, USA, pp. 1015-1018.

[23] Muhammad G and Alghathbar K. Environment Recognition for Digital Audio Forensics Using MPEG-7 and Mel Cepstral Features. Int Arab J Inf Technol 2013; 10(1).

[24] Kraetzer C, Oermann A, Dittmann J, Lang A. Digital audio forensics: a first practical evaluation on microphone and environment classification. In: 9th workshop on Multimedia & security; 2007; pp.63-74.

[25] Oermann A, Lang A and Dittmann J. Verifier-tuple for audio-forensic to determine speaker environment. In: 7th workshop on Multimedia and security; 2005; pp.57-62.

[26] Serizel R, Bisot V, Essid S, Richard G. Machine listening techniques as a complement to video image analysis in forensics. In: IEEE International Conference on Image Processing; 2016; IEEE, pp. 948-952.

[27] Güner A, Alçin ÖF Şengür A. Automatic digital modulation classification using extreme learning machine with local binary pattern histogram features. Measurement 2019;145: 214-225.

[28] Gupta D, Arora J, Agrawal U, Khanna A, Albuquerque VHC. Optimized Binary Bat algorithm for classification of white blood cells. Measurement 2019; 143:180-190.

[29] Hu Q, Ohata EF, Silva FH, Ramalho GL, Han T, Reboucas Filho PP. A new online approach for classification of pumps vibration patterns based on intelligent IoT system. Measurement 2020;151: 107138.

[30] Ma C, An Y, Shen E, Yu D, Zhang J. Car Interior Sound Field Zoning Using Optimal Loudspeaker Array and Double Iteration Method. J Audio Eng Soc 2024; 72(4): p. 247-256.

[31] Zaman K, Sah M, Direkoglu C, Unoki M. A survey of audio classification using deep learning. IEEE Access 202;11: 106620-106649.

[32] Li F, Pang X, Yang Z. Motor current signal analysis using deep neural networks for planetary gear fault diagnosis. Measurement 2019; 145:45-54.

[33] Xue Y, Dou D, Yang J, Multi-fault diagnosis of rotating machinery based on deep convolution neural network and support vector machine. Measurement 2020; 156:107571.

[34] Lu Y, Wang M, Wu W, Han Y, Zhang Q, Chen S, Dynamic entropy-based pattern learning to identify emotions from EEG signals across individuals. Measurement 2020; 150: 107003.

[35] Sudhagar S, Sakthivel M, Ganeshkumar P. Monitoring of friction stir welding based on vision system coupled with Machine learning algorithm. Measurement 2019; 144:135-143.

[36] Navarro JM and Pita A. Machine Learning Prediction of the Long-Term Environmental Acoustic Pattern of a City Location Using Short-Term Sound Pressure Level Measurements. Appl Sci 2023; 13(3):1613.

[37] Mollah A, Mahanta TK, Balide V., Intelligent Classification of Automotive Horn Sound Quality 2024; SAE Tech Pap, No: 2024-26-0204.

[38] Nasim F, Masood S, Jaffar A, Ahmad U, Rashid M. Intelligent Sound-Based Early Fault Detection System for Vehicles. Comput Syst Sci Eng 2023; 46(3).

[39] Fan X, Sun T, Chen W, Fan Q. Deep neural network based environment sound classification and its implementation on hearing aid app. Measurement 2020;107790.

[40] Jaber MM, Abd SK, Shakeel PM, Burhanuddin MA, Mohammed MA, Yussof S. A telemedicine tool framework for lung sounds classification using ensemble classifier algorithms. Measurement 2020; 107883.

[41] Shen G, Nguyen Q, Choi J, An environmental sound source classification system based on Mel-frequency cepstral coefficients and Gaussian mixture models. IFAC Proc Vol 2012; 45(6): 1802-1807.

[42] Saki F, Kehtarnavaz N. Real-time hierarchical classification of sound signals for hearing improvement devices. Appl Acoust 2018;132: 26-32.

[43] Salamon J, Jacoby C, Bello JP. A dataset and taxonomy for urban sound research. In: 22nd ACM international conference on Multimedia 2014;1041-1044.

[44] Medhat F, Chesmore D, Robinson J, Masked Conditional Neural Networks for sound classification. Appl Soft Comput 2020; 90: 106073.

[45] Chen Y, Guo Q, Liang X, Wang J, Qian Y. Environmental sound classification with dilated convolutions. Appl Acoust 2019;148:123-132.

[46] Souli S, Lachiri Z. Audio sounds classification using scattering features and support vectors machines for medical surveillance. Appl Acoust 2018;130: 270-282.

[47] López-Pacheco MG, Sánchez-Fernández LP, Molina-Lozano H, Sánchez-Pérez LA. Predominant environmental noise classification over sound mixing based on source-specific dictionary. Appl Acoust 2016; 112: 171-180.

[48] Tuncer T, Dogan S, Ertam F. Automatic voice based disease detection method using one dimensional local binary pattern feature extraction network. Appl Acoust 2019; 155: 500-506.

[49] AlQahtani MO. Environment Sound recognition for digital audio forensics using linear predictive coding features. In: International Conference on Digital Information Processing and Communications. 2011; Berlin, Heidelberg: Springer Berlin Heidelberg, pp.301-309.

[50] Black J. Youtube Channel. [cited 05.04.2024; Available from: https://www.youtube.com/channel/UCRT3s4cWpEKyu2U9xPtmOiw.

[51] Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KR. A fast iterative nearest point algorithm for support vector machine classifier design. IEEE Trans Neural Netw Learn Syst 2000; 11(1): 124-136.

[52] Kar S, Sharma KD, Maitra M, Gene selection from microarray gene expression data for classification of cancer subgroups employing PSO and adaptive K-nearest neighborhood technique. Expert Syst Appl 2015; 42(1):612-627.

[53] Liao Y, Vemuri VR, Use of k-nearest neighbor classifier for intrusion detection. Comput Secur 2002; 21(5): 439-448.

[54] Patel JR, Patel JM, Medical image fusion technique using singlelevel and multilevel DWT, Int J Eng Res Technol ESRSA Publications 2014; 3(1).

[55] Hasan KK, Ngah UK, Salleh MFM. Multilevel decomposition Discrete Wavelet Transform for hardware image compression architectures applications. In: 2013 IEEE International Conference on Control System, Computing and Engineering 2013;pp.315-320.

[56] Tian H and Ji W. A digital video watermarking scheme based on 1D-DWT. In: 2013 IEEE International Conference on Control System, Computing and Engineering IEEE, 2013;pp.315-320.

[57] Ojala T, Pietikäinen M, Mäenpää T. A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification. In: Advances in Pattern Recognition—ICAPR 2001: Second International Conference; Rio de Janeiro, Brazil; March 11–14, 2001; 2. Springer Berlin Heidelberg, 2001; pp. 399-408.

[58] Sun Y, Lou X, Bao B, A novel relief feature selection algorithm based on mean-variance model. J Inf Comput Sci 2011; 8(16): 3921-3929.

[59] Robnik-Šikonja M and Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. Mach Learn. 2003;53(1-2):23-69.

[60] Tuncer T, Dogan S, Pławiak P, Acharya UR. Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals. Knowl Based Syst 2019;186:104923.