Automation in Construction 125 (2021) 103645

Contents lists available at ScienceDirect

Automation in Construction

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

An automated location detection method in multi-storey buildings using environmental sound classification based on a new center symmetric nonlinear pattern: CS-LBlock-Pat

Mark Okaba, Turker Tuncer *

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

ARTICLE INFO

Keywords:
3D location detection in multi-storey buildings
Acoustical floor identification
Environmental sound classification
Center symmetric LBlock pattern
Statistical feature extraction
Iterative neighborhood component analysis

ABSTRACT

Classical navigations applications only give latitude and longitude information, thus, location detection has been processed in two-dimensional(2D) space by using latitude and longitude. There are many multi-storey buildings in cities, and these buildings cause location detection problem in 3D space. This research presents a solution for solving this problem. An environmental sound classification(ESC) dataset was collected from a multi-storey hospital, and an automated ESC model is presented. As a feature generation function, a new center symmetric nonlinear pattern is presented using a substitution box(S-Box) of LBlock lightweight block cipher., therefore, this model is named CS-LBlock-Pat. The presented model is applied to the dataset collected from a multi-storey hospital. This hospital has *ten* floors. Therefore, *ten* classes of classification results are given. This model yielded 95.38% accuracy rate using SVM. The results obtained from the classification phase obviously demonstrated that the 3D location detection problem could be solved by using ESC.

1. Introduction

Machine learning is very efficient and variable models have been presented to solve variable problems for instance computer vision [1–4], signal/image processing [5–7], sound recognition [8,9] and noise classification [10,11]. ESC is one of the topics of current interest in machine learning. The main objective of ESC is to detect the environment by using noises [12,13]. The human auditory system (HAS) cannot distinguish noises (ambient acoustic) since it has a limited sound processing capability. Computer-aided sound processing models have been used to overcome this problem [14].

ESC is a complex problem due to factors such as the overlap of different sounds, dimensionality, degree of noise, similarity, and different sound sources for signal processing models. Machine learning and deep learning models have been used to accurately predict location by using ESC [15,16]. The hand-designed features based machine learning models have generally used sound descriptors. These hand-designed features can reduce the noise and dimensionality of the signal, thus, more useful features are obtained from a sound signal. Traditional ESC methods concentrate on feature extraction and feature classification steps to achieve these effective features [17]. Methods such as Mel frequency cepstral coefficients (MFCC) [18,19], discrete wavelet transform (DWT) [20], short-time Fourier transform (STFT) [21,22] are widely used in the literature for feature extraction. Shallow classifiers, which KNN [23,24] and SVM [25,26] are preferred during the classification phase. However, the prior presented hand-crafted models have limited classification capabilities. Therefore, deep methods have been widely preferred in the ESC method to yield high classification performance. The sound signals are one-dimensional, hence, recurrent neural networks (RNNs) [27,28] have been employed to ESC datasets. However, parameter optimization is tough for RNNs models. Spectrogram based models have been used to simplify this process. In these models, the spectrogram image of the sound is extracted, and the extracted image is forwarded to convolutional neural networks (CNNs) [21,29,30]. This model can achieve higher classification rates. However, a specific model is not used for ESC. This research focuses on presenting a particular model for ESC. The high accurate ESC models have wide usage areas such as audio forensics, information security, public safety, smart cities, smart house applications, automation in construction, and location detection [25,31–35].

1.1. Motivation

The primary motivation of this paper is to solve the 3D location

* Corresponding author.
E-mail addresses: 191144119@firat.edu.tr (M. Okaba), turkertuncer@firat.edu.tr (T. Tuncer).

https://doi.org/10.1016/j.autcon.2021.103645
Received 15 July 2020; Received in revised form 14 February 2021; Accepted 21 February 2021
Available online 27 February 2021

detection problem by using an automated ESC method. Moreover, the living residential/multi-storey buildings have unique acoustic features. This research aims to detect the acoustical identity of the living residential buildings like hospitals, schools, or towers. Therefore, the acoustical identity of a multi-storey hospital is investigated. The widely used factors to evaluate buildings are design, engineering, construction technologies, and maintenance and management of constructed facilities. Several papers have used these factors but there is no paper about the human factors. However, the most important factor of a living building is the human factor in view of acoustic. Moreover, the architectures of most hospitals are similar in Turkey. Therefore, a human factor-based ESC model is presented. As known from the literature, the ESC model uses human factors. For instance, there are many kid noises in a kid garden and airports have unique acoustics. We analyzed acoustical features of each flat of a multi-storey building using the ESC methodology. This research aims to demonstrate the distinctive acoustical property of each floor of a living multi-storey building. As stated in the literature review, ESC has a wide usage area, including machine learning, information security, and digital forensics. 3D location detection is crucial for criminal tracking in digital forensics, and it is one of the problems of sound forensics, but there are limited works available on sound forensics. A new ESC dataset was collected from a multi-storey hospital to fill this gap, and a novel nonlinear pattern-based ESC model was presented. Our other motivation is to demonstrate the success of the S-boxes on the feature generation. Microstructures are very important hand-crafted feature extractors, and they generally use linear patterns. This work, however, tests performance in a nonlinear pattern, and this pattern is presented by using an S-box of the LBlock lightweight block cipher [36,37].

The work proposes a new hand-crafted and multilevel feature generation network. This network uses a combination of the proposed CS-LBlock-Pat, TQWT [38,39], and statistical feature generation. Furthermore, decomposition is processed by employing TQWT. TQWT is utilized for multilevel feature extraction, and 20 decomposed signals are obtained. The novelty of the presented multilevel feature generation network is the proposed CS- LBlock-Pat. This pattern is utilized as a primary textural feature extraction function with statistical moments supported by this feature generation network. 300 features are generated at each level., hence creating 6300 features in total (feature generation is applied on 20 decomposed sounds and raw environmental sound, $21 \times 300 = 6300$). In the second phase, the INCA [40] feature selector is employed for the extracted features. To classify these features selected LD [41], and SVM [42] classifiers are utilized in the last phase.

1.2. Contributions and novelties

Novelties and contributions of our method are given below.

Novelties;

- To detect 3D locations, a new ESC model is presented.
- A new ESC dataset was collected, and this dataset has been published publicly.
- A new generation feature generation function is presented, and this feature generator uses an S-box of a lightweight block cipher. Therefore, this feature generation function (CS- LBlock-Pat) is considered as a nonlinear pattern.

Contributions;

- This research aims to denote acoustical attributes of the living multi-storey buildings like hospital, courthouse. Therefore, a new dataset was collected from a multi-storey hospital. By using acoustical features of these building, each floor is detected easily. Moreover, this paper proposes a new sub-branch of the ESC.
- As stated, deep networks, multilevel, or multilayer feature generation models have high classification ability. Therefore, a novel generation feature generation network is presented by using TQWT. Statistical and textural feature generation functions are also used in this network.
- A high accuracy ESC model is presented to detect the third dimension of the location.

2. Literature review

Analyzing environmental sounds with ESC is a widely studied subject in the literature. Table 1 contains some of the recent studies with ESC.

3. Dataset

We collected a new ESC dataset for the detection of the third dimension location of a person. This research is focused on multi-storey buildings. Environmental sounds were collected from each of the *ten* floors of the Firat University Hospital on weekdays. We selected the hospital because it is a residential building. The collected sounds were segmented into a frame with a length of 48,000 since the sampling rate of the sounds collected is 48 kHz. The sounds gathered were collected from a mobile phone, a Huawei Mate 20 Lite device, and a file format of m4a. Researchers and developers can download this dataset from http://web.firat.edu.tr/turkertuncer/hospital.rar. There are *ten* classes in this dataset, and these classes represent each floor. Moreover, this dataset segmented into one-second segments, and the mat file format was uploaded on the web. The properties of this dataset are listed in Table 2.

4. The presented center symmetric LBlock pattern

This work presents a new generation nonlinear hand-designed feature generation function. To propose a nonlinear pattern, we utilized S-box. As can be seen in literature, the cryptography based feature generators have achieved high accuracies to solve one-dimensional signal classification problems [60,61]. As stated from block ciphers, S-boxes are nonlinear structures. Therefore, we selected an S-box since its nonlinearity has been proven in cryptology. Lightweight block ciphers are very popular, and they generally have used 4-bit S-boxes. By using these S-boxes, new generation nonlinear microstructures can be presented. This paper selects an S-box of the LBlock cipher to create a nonlinear pattern. The selected S-box is shown in Table 3 [36,37].

The presented CS-Lblock-Pat extracts binary features using this pattern (see Fig. 2) and the signum function (basic comparison function). The binary features extracted are converted to the decimal value to generate a map/feature signal. The histogram of this map/feature signal is utilized as features. Steps of the presented CS-Lblock-Pat are given below to better express this feature generation function.

0: Load signal/sound.

1: Divide signal/sound into 16 overlapping sized blocks.

$$block^{i} = sound(j+i-1), i = \{1, 2, \ldots, lngt-15\}, j = \{1, 2, \ldots, 16\} \tag{1}$$

where *block*, *sound* and *lngt* are overlapping block with a size of 16, input one-dimensional signal and length of this signal respectively. Eq. (1) defines overlapping block division using 16 sized blocks mathematically.

2: Generate binary features by using a pattern that is demonstrated in Fig. 1 and signum function

$$bt^{i}(k) = signum\left(block^{i}(S(k)), block^{i}(S(17-k))\right), k = \{1, 2, \ldots, 8\} \tag{2}$$

$$signum(x, y) = \begin{cases} 0, x-y<0 \\ 1, x-y \geq 0 \end{cases} \tag{3}$$

where bt^{i} are the bits extracted from ith block, *signum*(.,.) represents signum function and *x*, *y* are input parameters of the signum function, *S* (.) represents the used LBlock S-box which is defined in Table 3. Eqs. (2)–

Table 1
Literature review on ESC.

Studies	Year	Method	Database	Class	The accuracy result (%)
Zhang et al. method [43]	2019	Convolutional recurrent neural network	ESC-50 [44] DCASE2016 [45] ESC-10 [44] datasets	ESC-10 = 10 classes ESC-50 = 50 classes DCASE2016 = 15 classes	ESC-10 = 94.20 ESC-50 = 86.50 DCASE2016 = 88.90
Chi et al. method [46]	2019	Deep convolutional neural network	UrbanSound8K [47] and ESC-50 [44]	ESC-50 = 50 classes UrbanSound8K = 10 classes	UrbanSound8K = 80.30 ESC-50 = 83.80
Chandrakala and Jayalakshmi method [14]	2020	Mel frequency cepstral coefficients	DCASE2013 [45] DCASE2016 [45] ESC-10 [44] UrbanSound8K [47]	DCASE2013 = 10 classes DCASE2016 = 15 classes ESC-10 = 10 classes UrbanSound8K = 10 classes	DCASE2013 = 90.00 DCASE2016 = 85.90 ESC-10 = 74.00 UrbanSound8K = 85.47
Akbal method [48]	2020	One dimensional local binary pattern and ternary pattern	ESC-10 [44] dataset	10 classes	90.25
Dogan et al. method [49]	2020	Linear hexadecimal pattern	Collected data	25 classes	100.0
Demir et al. method [50]	2020	Convolutional Neural Networks	UrbanSound8K [47] and DCASE-2017 ASC [45]	UrbanSound8K = 10 classes DCASE-2017 ASC = 15 classes	UrbanSound8K = 86.70 DCASE-2017 ASC = 96.20
Park and Yoo method [51]	2020	One-dimensional convolutional neural network	UrbanSound8K [47] and ESC-50 [44]	ESC-50 = 50 classes UrbanSound8K = 10 classes	UrbanSound8K = 85.80 ESC-50 = 88.10
Ahmad et al. method [52]	2020	Empirical mode method	ESC-10 [44] dataset	10 classes	87.25
Guzhov et al. method [53]	2020	Residual Networks	ESC-50 [44] UrbanSound8K [47] ESC-10 [44] datasets	ESC-10 = 10 classes ESC-50 = 50 classes UrbanSound8K = 10 classes	ESC-10 = 97.00 ESC-50 = 91.50 UrbanSound8K mono = 84.20 stereo = 85.40
Su et al. method [54]	2020	Mel-Frequency Cepstral Coefficients and Log-mel Spectrogram	UrbanSound8K [47] and ESC-50 [44]	ESC-50 = 50 classes UrbanSound8K = 10 classes	UrbanSound8K = 93.40 ESC-50 = 85.60
Ullo et al. method [55]	2020	Convolutional neural network	ESC-10 [44] dataset	10 classes	95.80
Mushtaq et al. [56]	2021	Convolutional neural network	ESC-10 [44] ESC-50 [44], UrbanSound8K [47]	ESC-10 = 10 classes ESC-50 = 50 classes UrbanSound8K = 10 classes	ESC-10 = 99.04 ESC-50 = 97.57 UrbanSound8K = 99.49
Luz et al. [57]	2021	Convolutional neural network	ESC-50 [44] UrbanSound8K [47]	ESC-50 = 50 classes UrbanSound8K = 10 classes	ESC-50 = 86.20 UrbanSound8K = 96.16
Chandrakala et al. [58]	2021	Mel-frequency cepstral coefficients, Histogram of oriented gradients	ESC-50 [44] DCASE2016 [45] DCASE2018 [59]	ESC-50 = 50 classes DCASE2016 = 11 classes DCASE2018 = 41 classes	ESC-50 = 72.70 DCASE2016 = 91.30 DCASE2018 = 80.17

This research presents a new sub-branch of the ESC. As far as we know, it is the first work about the environmental sound classification of a multi-storey building. Therefore, ESC works are given in Table 1. As can be seen in Table 1, many studies/corpora have been presented to classify environmental sounds for detecting location. ESC is a complex issue for machine learning, thus, deep learning models have been used to overcome low prediction rate problem. More models have used convolutional neural network (CNN). These models converted sounds to images and they trained the generated images using a CNN. However, CNN models have high time burden and they are designed for computer vision. Therefore, an efficient machine learning model is presented for ESC in this research.

Table 2
Attributes of the speech emotion datasets used.

Class	*Number of SOUND environmental sound*	*Class*	*Number of SOUND environmental sound*
1	326	6	355
2	330	7	346
3	317	8	397
4	315	9	317
5	302	10	305

The graphical demonstrations of the categories are shown in Fig. 1.

(3) define the bit generation of the presented center symmetric and nonlinear feature generator. By deploying these equations, eight bits are generated from each overlapping block with a length of 16.

3: Convert the bits generated (*bt*) to decimal value for creating a map signal (*map*). This process is defined in Eq. (4).

$$map(i) = \sum_{k=1}^{8} bt^{i}(k)^{*}2^{k-1} \tag{4}$$

By employing Eq. (4), the generated eight bits are converted to a decimal value for producing the map signal.

4: Extract the histogram of the map signal. The map signal generated is coded with an 8-bit. Therefore, the length of the histogram extracted is calculated as $2^8 = 256$.

Table 3
The used S-box of the LBlock cipher.

x	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
S(x)	15	10	16	1	14	5	11	12	2	3	9	4	8	7	13	6

By using the S-box shown in Table 3, a center symmetric pattern is created. The pattern created by the S-box is shown in Fig. 2.

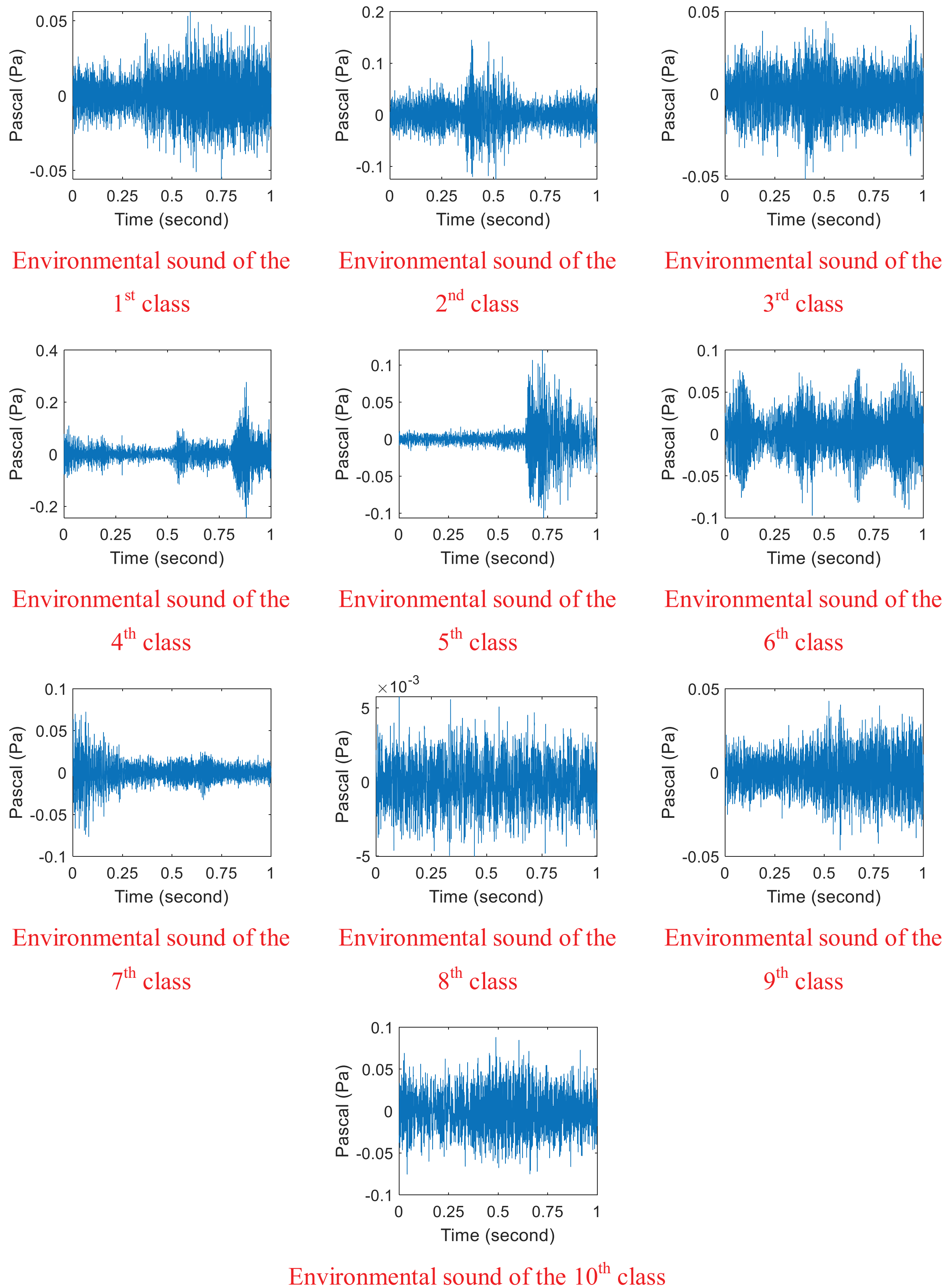

Fig. 1. Graphical demonstration of the collected dataset using sample environmental sounds.

$$histo(t) = 0; t = \{1, 2, \dots, 256\} \tag{5}$$

$$histo(map(i)) = histo(map(i)) + 1 \tag{6}$$

In Eq. (5), the initial value assignation of the histogram (*histo*) is showed. Histogram extraction is defined mathematically in Eq. (6).

The histogram extracted is utilized as a feature vector with a size of 256. The steps are given above (see steps 0–4) are defined as the feature generation procedure of the presented CV-LBlock-Pat. A numerical

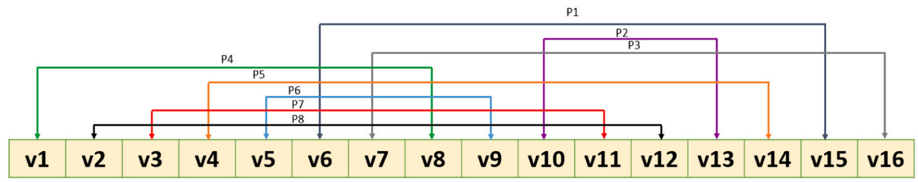

Fig. 2. The pattern of the proposed CV-Lblock-Pat.

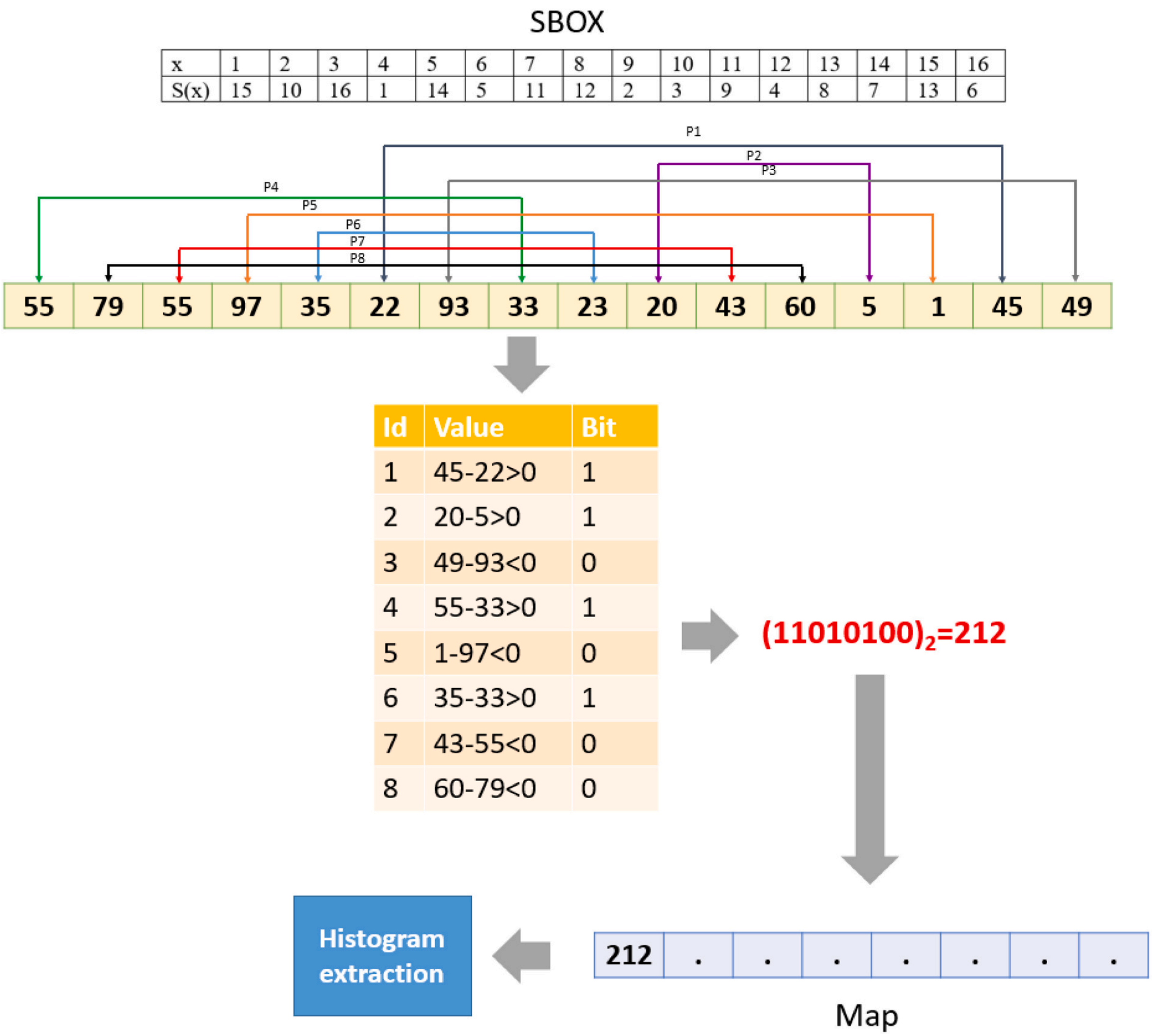

Fig. 3. Graphical summarization of the proposed CS-Lblock-Pat using a numerical exampler.

example of the presented CS-LBlock-Pat is shown in Fig. 3 for a better explanation.

5. The proposed automated floor detection model using environment sound classification

This work presents a new ESC model to detect floors in a multi-storey building using the proposed CS-LBlock-Pat and INCA based ESC model. As stated in the literature review, ESC is one of the current and complex research issues for machine learning and security. The primary objective of this model is to detect the 3D location of a suspect using ESC. The proposed model has three primary phases, namely, feature generation using the proposed CS-LBlock-Pat and statistical moments, informative/discriminative feature selection using INCA selector, and classification using two shallow classifiers.

A decomposition model is used to generate a feature in multilevel. In this step, TQWT is selected as a decomposition model. By using the proposed CS-LBlock-Pat and 22 statistical moments, 300 features are generated from a signal. In this network, the feature generation is processed from a raw sound signal and 20 decomposed signal. Therefore, 6300 features are generated from a sound. INCA, an improved version of the NCA selector was presented to solve the optimal number of features selection problem of the NCA. Therefore, INCA is selected as a feature selector. To better express, the success of the presented multilevel CS-LBlock-Pat based feature generation and INCA feature selection methods, two shallow/conventional classifiers are used to obtain results. An overview of the steps in this model is given below.

Step 0: Load ESC sound.

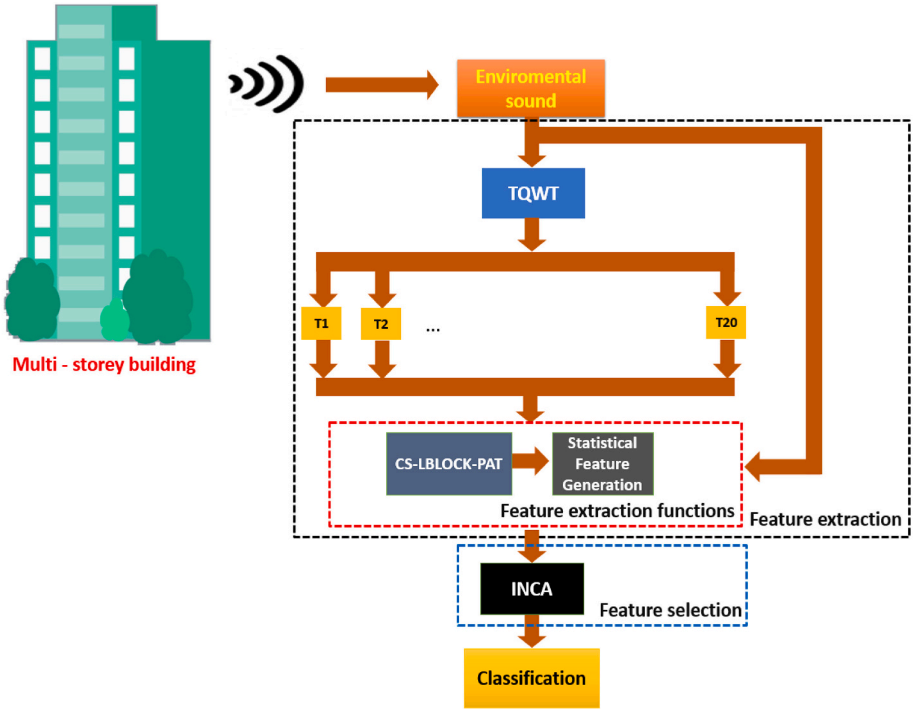

Fig. 4. Summarization of the proposed ESC model.

Step 1: Apply TQWT to sound signal collected by using 1, 3, and 19 for Q-factor (Q), redundancy (r), and the number of levels (j) parameters, respectively. 20 decomposed sound by using the given parameters.

Step 2: Generate features by using the proposed 22 statistical moments and the proposed CS-LBlock-Pat. 300 features are generated at each level. 256 of them are generated using the proposed CS-LBlock-Pat, 22 of them are generated applying statistical moments to signal (sound or decomposed sound), and 22 of them are generated by applying statistical moments of nonlinear textural features (CS-LBlock-Pat features).

Step 3: Concatenated the features extracted and obtain the final feature vector with a size of 21 x 300 = 6300.

Step 4: Select the most informative feature by employing INCA selector to the final feature vector. Here, INCA selected 523 features as the most informative for this problem.

Step 5: The feature vector selected by INCA are forwarded to conventional classifiers.

To illustratively explain this model, the block diagram of this model is shown in Fig. 4.

The details of the showed feature generation, selection, and classification phases are explained in subsections.

5.1. Feature generation

In this section, the details of the presented multilevel and hand-crafted feature generation function are given. As can be seen from Fig. 3, this model uses two feature generation strategy these are statistical and textural feature generations. Therefore, the proposed feature generation network is explained in three sections. These sections are textural feature generation, statistical feature generation, and general steps of the proposed multilevel feature generation network.

5.1.1. Textural feature generation

To generate textural features, the proposed CS-LBlock-Pat is utilized as a feature generation function, and details of this model are given in Section 3. To simply express this method, $CS-Lblock-Pat(.)$ function is used in general steps.

5.1.2. Statistical feature generation

One of the most used hand-crafted feature extraction function is statistical moments. The prime objective of this multilevel feature generation network is to generate valuable features by using both textural and statistical feature generation functions. Here, 22 statistical moments are used, and their equations are given below.

$$bf(1) = \frac{\sum_{i=1}^{L} sound_i}{L} \tag{7}$$

$$bf(2) = \sqrt{\frac{\sum_{i=1}^{L} (sound_i - bf(1))^2}{L}} \tag{8}$$

$$bf(3) = \sqrt{\frac{\sum_{i=1}^{L} (sound_i)^2}{L}} \tag{9}$$

$$bf(4) = -\sum_{i=1}^{L} prob(sound_i)*log(prob(sound_i)) \tag{10}$$

$$bf(5) = \sum_{i=1}^{L} \frac{i*sound_i - b(1)}{b(1)} \tag{11}$$

$$bf(6) = \frac{\sum_{i=1}^{L} \left| signal_{i+1} - signal_i \right|}{L} \tag{12}$$

$$bf(7) = \frac{\sqrt{L(L-1)}}{L-2} \left(\frac{\frac{1}{L}\sum_{i=1}^{L} (sound_i - b(1))^3}{\frac{1}{L}\sum_{i=1}^{L} (sound_i - b(1))^2} \right) \tag{13}$$

Procedure: $INCA(X, target, C)$

Input: Extracted features (X) with size of d x K, target with size of d and classifier (C).

Output: Final feature vector ($last$) with size of j.

00: Load X

01: for i=1 to K do

02: $X(:,i) = \frac{X(:,i)-min(X(:,i))}{max(X(:,i))-min(X(:,i))}$; //Apply min-max normalization to each column of the feature vector. NCA is a distance based feature selector. In order to use NCA effectively, normalization must be performed.

03: end for i

04: $endex = NCA(X, target)$; // Calculate indices of the features by using NCA. These indices are ordered by descending.

05: for k=1 to 900 do // Apply iterative NCA. Select number of features from 100 to 1000. The main aim of this step is to reduce computational complexity of this method.

06: for l=1 to 99+k do

07: $f^{NCA}(:,l) = X(:, endex(l))$; // Iterative NCA feature selection process.

08: end for l

09: $error(k) = C(f^{NCA})$; // Calculate error rates of each selected feature vector. In this work, kNN is used with 10-fold cross validation.

10: end for k

11: $[minimum, index] = \min(error)$; // Find minimum error and index of it.

12: for j=1 to $index + 99$ do // Select optimal features.

13: $last(:,j) = X(:, endex(j))$;

14: end for j

Fig. 5. Pseudocode of the INCA selector. As seen from this figure, INCA takes three parameters, namely feature vector, target, and classifier. NCA is a distance-based selector. Therefore, min-max normalization is applied to features, and this process is shown in Lines 01–03. NCA is applied to the normalized feature vector, and sorted indices are calculated. INCA has high time complexity. A number of feature range is chosen [100,1000] to decrease the time complexity of the INCA. The error/loss value of each selected feature vector is calculated using a classifier. Here, LD (C is LD) is employed as an error/loss value calculator with 10-fold cross-validation. In the Lines 11–14, the optimal feature selection process is shown.

$$bf(8) = \frac{L-1}{(L-2)(L-3)}\left[(L+1)\left(\left(\frac{\frac{1}{L}\sum_{i=1}^{L}(sound_i - b(1))^4}{\frac{1}{L}\sum_{i=1}^{L}(sound_i - b(1))^2}\right) - 3\right) + 6\right] \tag{14}$$

$$bf(9) = sound\left(\frac{L}{2}\right) \tag{15}$$

$$bf(10) = min\{sound\} \tag{16}$$

$$bf(11) = max\{sound\} \tag{17}$$

$$bf(12) = \frac{bf(1)}{bf(2)} \tag{18}$$

$$bf(13) = max\{sound\} - median\{sound\} \tag{19}$$

$$bf(14) = max\{sound\}/median\{sound\} \tag{20}$$

$$bf(15) = max\{sound\}/b(1) \tag{21}$$

$$bf(16) = bf(1)/min\{sound\} \tag{22}$$

$$bf(17) = bf(1)/median\{sound\} \tag{23}$$

$$bf(18) = bf(1) - min\{sound\} \tag{24}$$

$$bf(19) = bf(1) - median\{sound\} \tag{25}$$

$$bf(20) = bf(2) - median\{sound\} \tag{26}$$

$$bf(21) = bf(2)/median\{sound\} \tag{27}$$

$$bf(21) = bf(2)/median\{sound\} \quad (28)$$

where bf represents extracted statistical features, $prob$ defines the probability of the sounds. The given equations (Eqs. (7)–(28)) are defined the used statistical features. They are widely used statistical moments to generate statistical features.

5.1.3. General steps of the proposed feature generation network

0: Load environmental sound.

1: Apply TQWT to sound signal and generate 20 sub-bands (SB).

$$SB = TQWT(sound, 1, 3, 19), Q = 1, r = 3, J = 19 \quad (29)$$

The used main decomposer is TQWT and three parameters are used to generate sub-bands (SB) of TQWT. These parameters are Q-factor (Q), redundancy (r) and level number (J). Q value is used to determine oscillatory value. By employing the given parameters in Eq. (29), 20 sub-bands are generated. Mid-level and high-level features are generated using these sub-bands and the used feature generation functions.

2: Extract fused features from the sound signal and sub-bands of the sound signal

$$feat^1 = concat(CS - Lblock - Pat(sound), statis(sound), statis(CS - Lblock - Pat(sound))) \quad (30)$$

$$feat^r = conc\left(CS - Lblock - Pat\left(SB^{r-1}\right), statis\left(SB^{r-1}\right), statis\left(CS - Lblock - Pat\left(SB^{r-1}\right)\right)\right) \quad (31)$$

$$r = \{2, 3, \ldots, 21\} \quad (32)$$

In Eqs. (30)–(32), $statis(.)$ and $CS - Lblock - Pat(.)$ represent the used statistical feature generation function and CS-LBlock-Pat nonlinear text respectively, $conc(.)$ defines concatenation function.

3: Concatenate the feature vectors generated with size 300 and obtain the last feature vector.

$$X = conc\left(feat^1, feat^2 \ldots feat^r\right) \quad (33)$$

where X is the last feature vector with a size of 6300.

5.2. Feature selection

An improved version of the NCA is utilized to select the most informative features from the 6300 extracted features as a feature selector and is named INCA. INCA was presented by Tuncer et al. [40] to select the optimum number of features automatically. To better express, the INCA feature selector, the NCA feature selection function is defined as

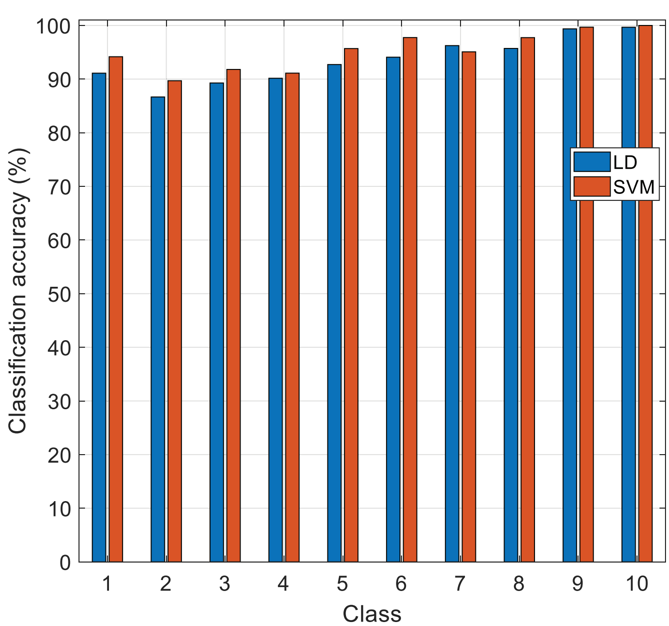

Fig. 6. The calculated accuracy rates for each class (floor) using LD and SVM classifier. The used Quadratic SVM reached higher classification rates for each class except for 7th class, and this classifier reached magnificent classification accuracy (100.0%) in the 10th class.

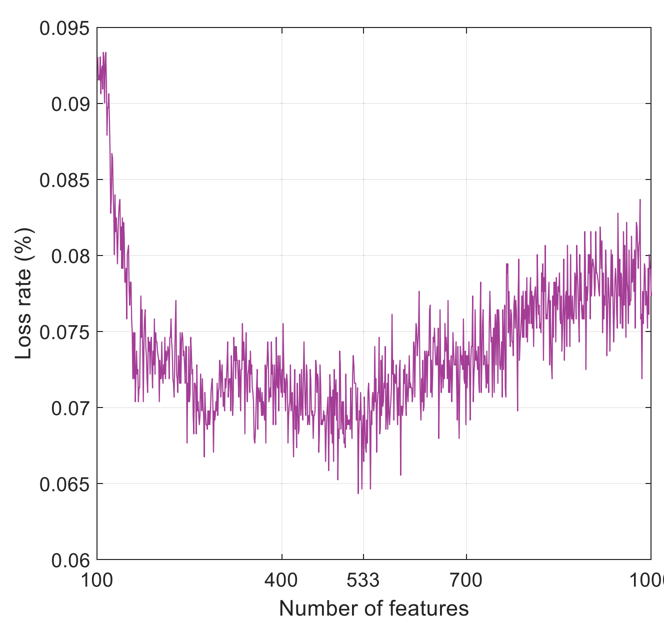

Fig. 7. The feature selection of the INCA. Here, LD is utilized as an error value calculator, since it has low computational complexity.

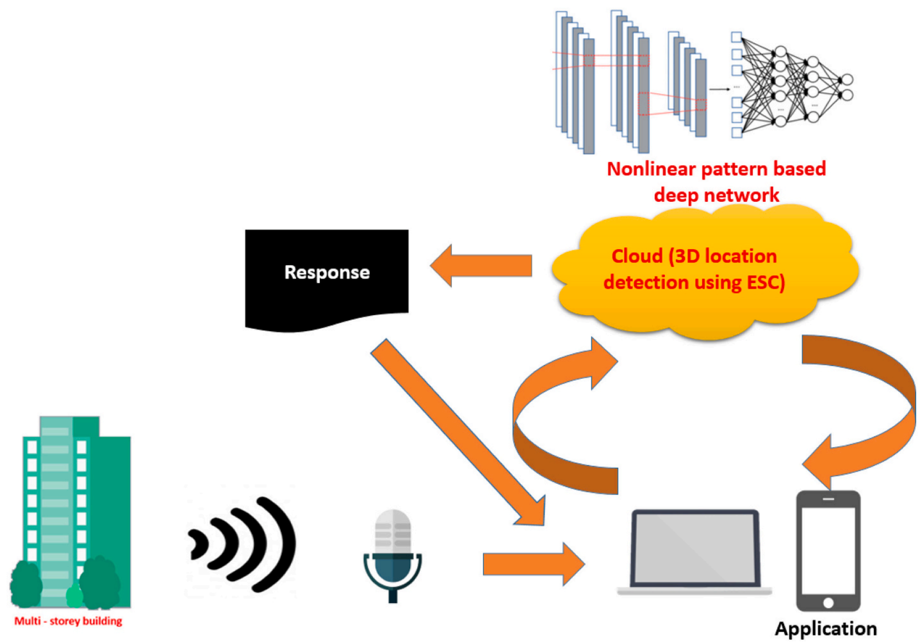

Fig. 8. Snapshot of the planned future work.

NCA(.). The procedure of the INCA selector is shown in Fig. 5.

The most informative 523 features are selected in this phase.

5.3. Classification

The last phase of the proposed method is classification, and *two* shallow classifiers are used for obtaining measurements. The used classifiers are LD and SVM classifiers. These classifiers are used by employing the MATLAB classification learner toolbox. 10-fold cross-validation is utilized for validation and testing. The attributes of the used classifiers are listed below.

LD: LD is one of the most known linear classifiers. This classifier has no parameter in the MATLAB classification learner. Therefore, no parameter was set [41].

SVM: SVM is one of the best shallow classifiers and uses variable kernels. This research uses Quadratic (2nd-degree polynomial kernel) SVM. C (box constraint level) value is selected as 1 and the multi-classification method is one-vs-all [42].

6. Performance analysis

6.1. Experimental setup

The proposed CS-LBlock-Pat and INCA based classification model was programmed by MATLAB (2020a) environment on a desktop computer with a straightforward configuration. No parallel programming or GPU core was used to implement this model. Firstly, the gathered sounds from each floor of the hospital (multi-storey building) were stored to create the dataset. They were labeled and segmented into one-second length. The prime objective of using short frames is to obtain a fast respond model. This model was coded using functions. The used function functions are CS-LBlock-Pat, statistical feature generation, and INCA. The classification learner (MATLAB) was used to classify the features selected.

Table 4
The mathematical equality of performance metrics.

Performance metric	Equations	No
cac	$\frac{tp+tn}{tp+tn+fp+fn}$	(34)
rec	$\frac{tp}{tp+fn}$	(35)
pre	$\frac{tp}{tp+fp}$	(36)
F1	$\frac{2tp}{2tp+fn+fp}$	(37)

Table 5
The calculated results per each used classifier.

Classifier	Performance metric	Performance rate (%)
LD	Classification accuracy	93.53
	Average recall	93.50
	Average precision	93.58
	F1	93.54
SVM	Classification accuracy	95.32
	Average recall	95.27
	Average precision	95.26
	F1	95.27

6.2. Results

This model used *two* shallow classifiers to obtain results. In the literature, variable measurements have been used to test the performance of a classifier. In this work, performance metrics accuracy (*cac*), recall (*rec*), precision (*pre*), and F1-score (*F1*) are calculated using a number of true positives (*tp*), true negatives (*tn*), false positives (*fp*) and false negative (*fn*) [62].

In this section, the performance parameters are listed in Table 4 and the obtained results are given in Table 5.

The confusion matrices of the classifiers used (LD and SVM) are listed to comprehensively demonstrate the obtained results in Tables 6 and 7.

The calculated accuracy values for each floor per classifier is shown in Fig. 6.

7. Discussions

The primary aim of this paper is to detect the 3rd dimension of location using ESC. Therefore, a dataset was gathered from Firat University Hospital. We segmented each sound in the one-second sized frame for proposing a fast respond system. Moreover, we have presented

Table 6
Confusion matrix of the LD classifier.

True class	Predicted class									
	1	2	3	4	5	6	7	8	9	10
1	297	19	5	4	0	0	1	0	0	0
2	28	286	4	12	0	0	0	0	0	0
3	7	9	283	12	0	3	3	0	0	0
4	15	10	3	284	1	0	0	2	0	0
5	1	0	2	1	280	3	13	1	1	0
6	2	1	3	3	3	334	8	1	0	0
7	1	1	5	2	2	2	333	0	0	0
8	1	0	0	6	6	4	0	380	0	0
9	0	0	1	0	0	1	0	0	315	0
10	0	0	0	0	0	0	1	0	0	304

Table 7
Confusion matrix of the SVM classifier.

True class	Predicted class									
	1	2	3	4	5	6	7	8	9	10
1	307	11	4	2	0	0	2	0	0	0
2	15	296	7	12	0	0	0	0	0	0
3	3	7	291	10	1	1	4	0	0	0
4	9	11	4	287	2	0	0	1	1	0
5	0	0	3	1	289	2	7	0	0	0
6	0	2	2	1	1	347	2	0	0	0
7	1	2	6	1	3	2	329	2	0	0
8	2	0	0	4	3	0	0	388	0	0
9	0	0	0	1	0	0	0	0	316	0
10	0	0	0	0	0	0	0	0	0	305

Table 8
Mathew's correlation coefficients (MCC) and Cohen's Kappa (CK) results (%) of the presented CS-LBlock pattern and INCA based floor detection model.

Classifier	MCC	CK
LD	92.81	92.81
SVM	94.75	94.79

a new nonlinear pattern by using one of the S-Box of a lightweight block cipher (LBlock cipher). This pattern has been created center symmetrically. Therefore, the presented model is named CS-Lblock-Pat. The prime objective of this pattern is to capture nonlinear features from a sound signal. The 256 features generated by the proposed CS-Lblock-Pat captured hidden nonlinearities successfully. Also, the statistical features were generated to enforce the proposed feature generation network. INCA was employed as a feature selector, and the feature selection process of the INCA is shown in Fig. 7.

Two shallow classifiers have been used in this research, and these are SVM and LD classifiers. These classifiers have been used to illustrate the success of the proposed feature generation and selection methods. Results clearly demonstrated that the SVM classifier achieved 95.38% accuracy. Moreover, Mathew's correlation coefficients and Cohen's Kappa are calculated and these results are listed in Table 8.

Benefits of the presented model are:

- Since the navigations return only latitude and longitude information. An ESC based 3D location detection is presented in this work, and it is the first ESC based solution for this problem. In this view, this research discovers an unknown property of the ESC.
- A dataset was gathered to solve the 3D location detection problem, and it was published publicly.
- A new pattern (CS-Lblock-Pat) is presented to find nonlinear attributes (hidden nonlinear pattern) of an environmental sound.
- The presented feature generation network is multilevel. Here, the effectiveness of hand-crafted methods was used together.
- A classification model is presented to detect the 3rd-dimension of the location with high performance.

The limitation of this work is as follows. More sounds can be collected from variable multi-storey buildings.

We intend to collect more sounds from a variable multi-storey building. Especially, there are many suspicious and multi-storey buildings in the metropolises. This model can be used in digital forensics and security applications. A new sound forensics toolbox can be developed using our presented model. By applying web services and deep learning, 3D location can be detected using our presented future model. Moreover, S-box or permutation box based new generation feature extractors can be presented, and new nonlinear patterns based deep learning networks can be presented. The snapshot of the intended framework is shown in Fig. 8.

8. Conclusion

This research presented a high accuracy ESC model using multilevel hand-crafted (nonlinear and statistical features) feature generation model and INCA feature selector. The prime aim of this model is to detect the 3D location of a person using the presented ESC model in a multi-storey building. This ESC model has three phases, namely feature generation using the proposed CS-LBlock-Pat and TQWT. The most informative features were selected by employing the INCA selector. The presented feature generation CS-LBlock-Pat and TQWT based feature generation network extract 6300 features from each sound, and INCA selected 523 most valuable features. These features were classified using LD and SVM classifiers, and these classifiers reached 93.53% and 95.32%, respectively. The results obtained clearly demonstrated the success of the presented CS-LBlock-Pat and INCA based ESC model for 3D location detection. The collected dataset was also published publicly for improving ESC based location detection works.

Declaration of Competing Interest

There is no 'Conflict of Interest' in the publication of the manuscript "*An automated location detection method in multi-storey buildings using environmental sound classification based on a new center symmetric nonlinear pattern: CS-LBlock-Pat*".

Acknowledgements

This research is supported by Firat University Scientific Project Research Fund, Turkey. Project number is TEKF.20.10.

References

[1] M. Toğaçar, B. Ergen, Z. Cömert, F. Özyurt, A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models, Irbm 41 (2020) 212–222.
[2] F. Özyurt, A fused CNN model for WBC detection with MRMR feature selection and extreme learning machine, Soft. Comput. 24 (2020) 8163–8172.
[3] F. Özyurt, Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures, J. Supercomput. (2019) 1–19.
[4] E. Aydemir, F. Kaysi, Development of short forms of scales with decision tree algorithms, Balkan J. Electric. Comput. Eng. 8 (2021) 285–290.
[5] E. Aydemir, Weka ile yapay zeka, Seçkin Yayınevi, Ankara, 2018.
[6] G. Ozmen Koca, S. Dogan, H. Yilmaz, A multi-objective route planning model based on genetic algorithm for cuboid surfaces, Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije 59 (2018) 120–130.
[7] T. Tuncer, S. Dogan, A. Subasi, A new fractal pattern feature generation function based emotion recognition method using EEG, Chaos, Solitons Fractals 144 (2021) 110671.
[8] S. Dogan, A new fractal H-tree pattern based gun model identification method using gunshot audios, Appl. Acoust. 177 (2021) 107916.
[9] T. Tuncer, E. Akbal, S. Dogan, Multileveled ternary pattern and iterative ReliefF based bird sound classification, Appl. Acoust. 176 (2021) 107866.
[10] S. Martina, S. Gherardini, F. Caruso, Machine Learning Approach for Quantum Non-Markovian Noise Classification, arXiv preprint arXiv:2101.03221, 2021.
[11] G. Algan, I. Ulusoy, Image classification with deep learning in the presence of noisy labels: a survey, Knowl.-Based Syst. 106771 (2021).
[12] J. Salamon, J.P. Bello, Deep convolutional neural networks and data augmentation for environmental sound classification, IEEE Sig. Proc. Lett. 24 (2017) 279–283.
[13] X. Fan, T. Sun, W. Chen, Q. Fan, Deep neural network based environment sound classification and its implementation on hearing aid app, Measurement 107790 (2020).
[14] S. Chandrakala, S. Jayalakshmi, Generative model driven representation learning in a hybrid framework for environmental audio scene and sound event recognition, IEEE Trans. Multimedia 22 (2019) 3–14.
[15] J.-C. Wang, C.-H. Lin, B.-W. Chen, M.-K. Tsai, Gabor-based nonuniform scale-frequency map for environmental sound classification in home automation, IEEE Trans. Autom. Sci. Eng. 11 (2013) 607–613.
[16] M. Jung, S. Chi, Human activity classification based on sound recognition and residual convolutional neural network, Autom. Constr. 114 (2020) 103177.
[17] M. Esmaeilpour, P. Cardinal, A.L. Koerich, Unsupervised feature learning for environmental sound classification using weighted cycle-consistent generative adversarial network, Appl. Soft Comput. 86 (2020) 105912.
[18] G. Shen, Q. Nguyen, J. Choi, An environmental sound source classification system based on Mel-frequency cepstral coefficients and Gaussian mixture models, IFAC Proc. Vol. 45 (2012) 1802–1807.
[19] S. Chu, S. Narayanan, C.-C.J. Kuo, Environmental sound recognition with time–frequency audio features, IEEE Trans. Audio Speech Lang. Process. 17 (2009) 1142–1158.
[20] T. Lenin, P. Dhanalakshmi, Environmental sound classification using discrete wavelet transform, Adv. Nat. Appl. Sci. 9 (2015) 539–545.
[21] M. Huzaifah, Comparison of Time-Frequency Representations for environmental sound classification using convolutional neural networks, arXiv preprint arXiv:1706.07156, 2017.
[22] B.-J. Han, E. Hwang, Environmental sound classification based on feature collaboration, in: 2009 IEEE International Conference on Multimedia and Expo, IEEE, 2009, pp. 542–545.
[23] J.-C. Wang, J.-F. Wang, K.W. He, C.-S. Hsu, Environmental sound classification using hybrid SVM/KNN classifier and MPEG-7 audio low-level descriptor, in: The 2006 IEEE International Joint Conference on neural Network Proceedings, IEEE, 2006, pp. 1731–1735.
[24] Y. Li, Y. Li, Eco-environmental sound classification based on matching pursuit and support vector machine, in: 2010 2nd International Conference on Information Engineering and Computer Science, IEEE, 2010, pp. 1–4.
[25] P. Khunarsal, C. Lursinsap, T. Raicharoen, Very short time environmental sound classification based on spectrogram pattern matching, Inf. Sci. 243 (2013) 57–74.
[26] Y.-T. Peng, C.-Y. Lin, M.-T. Sun, K.-C. Tsai, Healthcare audio event classification using hidden Markov models and hierarchical hidden Markov models, in: 2009 IEEE International Conference on Multimedia and Expo, IEEE, 2009, pp. 1218–1221.
[27] Z. Zhang, S. Xu, T. Qiao, S. Zhang, S. Cao, Attention based convolutional recurrent neural network for environmental sound classification, in: Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Springer, 2019, pp. 261–271.
[28] S. Sigtia, A.M. Stark, S. Krstulović, M.D. Plumbley, Automatic environmental sound recognition: performance versus computational cost, IEEE/ACM Trans. Audio Speech Lang. Process. 24 (2016) 2096–2107.
[29] K.J. Piczak, Environmental sound classification with convolutional neural networks, in: 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, 2015, pp. 1–6.
[30] Y. Su, K. Zhang, J. Wang, K. Madani, Environment sound classification using a two-stream CNN based on decision-level fusion, Sensors 19 (2019) 1733.
[31] G. Muhammad, K. Alghathbar, Environment recognition for digital audio forensics using MPEG-7 and Mel cepstral features, Int. Arab J. Inform. Technol. (IAJIT) 10 (2013).
[32] F. Su, L. Yang, T. Lu, G. Wang, Environmental sound classification for scene recognition using local discriminant bases and HMM, in: Proceedings of the 19th ACM International Conference on Multimedia, 2011, pp. 1389–1392.
[33] X. Wang, M.J. Kim, P.E. Love, S.-C. Kang, Augmented reality in built environment: classification and implications for future research, Autom. Constr. 32 (2013) 1–13.
[34] S. Moncrieff, S. Venkatesh, G. West, Dynamic privacy assessment in a smart house environment using multimodal sensing, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 5 (2008) 1–29.
[35] J.P. Bello, C. Mydlarz, J. Salamon, Sound analysis in smart cities, in: Computational Analysis of Sound Scenes and Events, Springer, 2018, pp. 373–397.
[36] W. Wu, L. Zhang, LBlock: A lightweight block cipher, in: International Conference on Applied Cryptography and Network Security, Springer, 2011, pp. 327–344.
[37] S. Sun, L. Hu, P. Wang, K. Qiao, X. Ma, L. Song, Automatic security evaluation and (related-key) differential characteristic search: Application to SIMON, PRESENT, LBlock, DES (L) and other bit-oriented block ciphers, in: International Conference on the Theory and Application of Cryptology and Information Security, Springer, 2014, pp. 158–178.
[38] A. Subasi, S.M. Qaisar, Surface EMG signal classification using TQWT, Bagging and Boosting for hand movement recognition, J. Ambient. Intell. Humaniz. Comput. (2020), https://doi.org/10.1007/s12652-020-01980-6.
[39] E. Aydemir, T. Tuncer, S. Dogan, A Tunable-Q wavelet transform and quadruple symmetric pattern based EEG signal classification method, Med. Hypotheses 134 (2020) 109519.
[40] T. Tuncer, S. Dogan, F. Özyurt, S.B. Belhaouari, H. Bensmail, Novel multi center and threshold ternary pattern based method for disease detection method using voice, IEEE Access 8 (2020) 84532–84540.
[41] M. Fraz, P. Remagnino, A. Hoppe, S. Barman, Retinal image analysis aimed at extraction of vascular structure using linear discriminant classifier, in: 2013 International Conference on Computer Medical Applications (ICCMA), IEEE, 2013, pp. 1–6.
[42] S.S. Keerthi, S.K. Shevade, C. Bhattacharyya, K.R. Murthy, A fast iterative nearest point algorithm for support vector machine classifier design, IEEE Trans. Neural Netw. 11 (2000) 124–136.
[43] Z. Zhang, S. Xu, S. Zhang, T. Qiao, S. Cao, Learning attentive representations for environmental sound classification, IEEE Access 7 (2019) 130327–130339.
[44] K.J. Piczak, ESC: Dataset for environmental sound classification, in: Proceedings of the 23rd ACM International Conference on Multimedia, 2015, pp. 1015–1018.
[45] A. Mesaros, T. Heittola, T. Virtanen, TUT database for acoustic scene classification and sound event detection, in: 2016 24th European Signal Processing Conference (EUSIPCO), IEEE, 2016, pp. 1128–1132.
[46] Z. Chi, Y. Li, C. Chen, Deep convolutional neural network combined with concatenated spectrogram for environmental sound classification, in: 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), IEEE, 2019, pp. 251–254.
[47] J. Salamon, C. Jacoby, J.P. Bello, A dataset and taxonomy for urban sound research, in: Proceedings of the 22nd ACM international conference on Multimedia, 2014, pp. 1041–1044.
[48] E. Akbal, An automated environmental sound classification methods based on statistical and textural feature, Appl. Acoust. 167 (2020) 107413.
[49] S. Dogan, E. Akbal, T. Tuncer, A novel ternary and signum kernelled linear hexadecimal pattern and hybrid feature selection based environmental sound classification method, Measurement (2020) 108151.
[50] F. Demir, D.A. Abdullah, A. Sengur, A new deep CNN model for environmental sound classification, IEEE Access 8 (2020) 66529–66537.
[51] H. Park, C.D. Yoo, CNN-based learnable gammatone filterbank and equal-loudness normalization for environmental sound classification, IEEE Sig. Proc. Lett. 27 (2020) 411–415.
[52] S. Ahmad, S. Agrawal, S. Joshi, S. Taran, V. Bajaj, F. Demir, A. Sengur, Environmental sound classification using optimum allocation sampling based empirical mode decomposition, Phys. A: Stat. Mech. Appl. 537 (2020) 122613.
[53] A. Guzhov, F. Raue, J. Hees, A. Dengel, ESResNet: Environmental Sound Classification Based on Visual Domain Models, arXiv preprint arXiv:2004.07301, 2020.
[54] Y. Su, K. Zhang, J. Wang, D. Zhou, K. Madani, Performance analysis of multiple aggregated acoustic features for environment sound classification, Appl. Acoust. 158 (2020) 107050.
[55] S.L. Ullo, S.K. Khare, V. Bajaj, G.R. Sinha, Hybrid computerized method for environmental sound classification, IEEE Access 8 (2020) 124055–124065.
[56] Z. Mushtaq, S.-F. Su, Q.-V. Tran, Spectral images based environmental sound classification using CNN with meaningful data augmentation, Appl. Acoust. 172 (2021) 107581.

[57] J.S. Luz, M.C. Oliveira, F.H. Araújo, D.M. Magalhães, Ensemble of handcrafted and deep features for urban sound classification, Appl. Acoust. 175 (2021) 107819.

[58] S. Chandrakala, M. Venkatraman, N. Shreyas, S. Jayalakshmi, Multi-view representation for sound event recognition, signal, Image Video Process. (2021) 1–9.

[59] E. Fonseca, M. Plakal, F. Font, D.P. Ellis, X. Favory, J. Pons, X. Serra, General-Purpose Tagging of Freesound Audio with Audioset Labels: Task Description, Dataset, and Baseline, arXiv preprint arXiv:1807.09902, 2018.

[60] T. Tuncer, S. Dogan, U.R. Acharya, Automated accurate speech emotion recognition system using twine shuffle pattern and iterative neighborhood component analysis techniques, Knowl.-Based Syst. 211 (2021) 106547.

[61] O. Yaman, T. Tuncer, B. Tasar, DES-Pat: a novel DES pattern-based propeller recognition method using underwater acoustical sounds, Appl. Acoust. 175 (2021) 107859.

[62] T. Tuncer, S. Dogan, P. Pławiak, U.R. Acharya, Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals, Knowl.-Based Syst. 186 (2019) 104923.