

# Vehicle Interior Sound Classification Based on Local Quintet Magnitude Pattern and Iterative Neighborhood Component Analysis

Erhan Akbal, Turker Tuncer & Sengul Dogan

To cite this article: Erhan Akbal, Turker Tuncer & Sengul Dogan (2022) Vehicle Interior Sound Classification Based on Local Quintet Magnitude Pattern and Iterative Neighborhood Component Analysis, Applied Artificial Intelligence, 36:1, 2137653, DOI: [10.1080/08839514.2022.2137653](https://doi.org/10.1080/08839514.2022.2137653)

To link to this article: <https://doi.org/10.1080/08839514.2022.2137653>



© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 20 Oct 2022.



Submit your article to this journal [↗](#)



Article views: 744



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

# Vehicle Interior Sound Classification Based on Local Quintet Magnitude Pattern and Iterative Neighborhood Component Analysis

Erhan Akbal , Turker Tuncer , and Sengul Dogan 

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

## ABSTRACT

Nowadays, environmental sound classification (ESC) has become one of the most studied research areas. Sound signals that are indistinguishable from the human auditory systems have been classified by computer-aided systems and machine learning methods. Therefore, ESC has been used in signal processing and sound forensics applications. A novel ESC type is presented in this paper, and it is named as vehicle interior sound classification (VISC). VISC is defined as one of the sub-branches of the ESC, and it is utilized as sound-based biometrics for vehicles. A hand-crafted feature-based VISC method is presented. The proposed method has multileveled feature generation by using maximum pooling and the proposed local quintet magnitude pattern (LQMP), feature selection with iterative neighborhood component analysis (INCA), and classification phases. A novel VISC dataset was collected from YouTube and the proposed LQMP and INCA based method applied to the collected sounds. The results denoted that following: the accuracy, F1-score, and geometric mean of the proposed LQMP and INCA based VISC method were calculated as 98.38%, 98.23%, and 98.21% by using support vector machine classifier respectively. The contribution of the proposed VISC method is to denote that the vehicles can be classified by using sound.

## ARTICLE HISTORY

Received 4 August 2020  
Revised 4 January 2022  
Accepted 13 October 2022

## Introduction

Environmental sound classification (ESC) has been used for different purposes in recent years and therefore it attracts the attention of researchers. In these studies, vehicle interior and exterior sounds (Hashimoto and Takao 1990; Västfjäll et al. 2002; Wang et al. 2013), internal organ sounds (Mishra, Menon, and Mukherjee 2018; Patidar and Pachori 2014; Randhawa and Singh 2015; Wang et al. 2013), fault detection from device sound (Gramatikov et al. 2016; Scanlon, Kavanagh, and Boland 2012), air conditioner and refrigerator noise (Jeon et al. 2011), aircraft and helicopter sounds (Akhtar, Elshafei-Abmed, and Ahmed 2001; More and Davies 2010), ambient sounds (Huang et al. 2019; Lurz et al. 2017), animal sounds (Ko et al. 2019), underwater monitoring (Mayer, Magno, and Benini 2019), gender

**CONTACT** Sengul Dogan  [sdogan@firat.edu.tr](mailto:sdogan@firat.edu.tr)  Department of Digital Forensics Engineering, Technology Faculty, Firat University, Elazig 23119, Turkey

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

recognition and classification using music (Chang, Chen, and Lee 2021; Liu et al. 2021), sound classification using acoustic properties (Nanni et al. 2017; Yang and Zhao 2021) have been used. ESC has limited information in terms of time and frequency properties. In addition, deterministic methods cannot achieve good performance since environmental sounds often differ according to the environment. Therefore, signal processing and machine learning techniques have been used in recent years.

Today, people have been mostly used their own vehicles for transportation ("ACEAReport Vehicles in use Europe 2019" 2019). Therefore, people spend a lot of time in vehicles (Boucher and Noyer 2012). Manufacturers have produced many types of land vehicles for different purposes, and these vehicle types have different characteristics. Design, engine sound, cabin interior sound, door closing sound, exhaust sound, headlight light of vehicles can be considered as biometric data of vehicles (Manders et al. 2000). Vehicle interior noise affects the preferences of users (Lemaitre and Susini 2019). Minimizing vehicle interior noise is a vital engineering process for manufacturers. There are also studies showing that the vehicle interior noise causes psychological effects during the trip (Takada et al. 2019).

Vehicles produce many different sounds during the drive. The sounds produced contain many descriptive data. The brand and model of the vehicle, engine failures, insulation level, cabin comfort level, material effect, exhaust measurement can be analyzed using sound signals. All sounds produced are considered as environmental sound. Ambient recognition or ESC has also been widely used for detecting criminal events. ESC has been used to detect the location of the crime scene (Al-Ali, Senadji, and Naik 2017; Khan et al. 2018). One of the sub-classes of ESC is vehicles. However, there is no special ESC method for vehicles. The studies have been generally proposed for ESC. Some of the related studies about cybercrime and vehicles are given as follows. Baucas and Spachos (Baucas and Spachos 2020) proposed a model for differentiating ambient sounds and urban sounds using the Mel-frequency cepstral coefficient and short-time Fourier transform feature extraction method. UrbanSound8K dataset (Salamon, Jacoby, and Bello 2014) was used in this study. Khan et al. (Khan et al. 2018) presented a method of detecting sounds from multimedia devices in IT crimes. The primary purpose of their study is to determine the microphone, acoustic environment, and scene. The recording, the codec, and the artifacts resulting from the ambient were detected by their detection method. Mazarakis and Avaritsiotis (Mazarakis and Avaritsiotis 2007) classified the data obtained from the Fiber Bragg Grating sensor system for classification of vehicles. In the dataset used (Duarte and Hu 2004), the model tested on Alexnet they achieved 87.1% classification accuracy rate. Wei et al. (Jie Wei et al. 2019) proposed a specialist model to remotely classify vehicle engines using a Laser Doppler Vibrometer (LDV). Artificial neural network and convolutional neural networks methods were chosen as classifiers in their study. Tone-pitch indexes were used and high accuracy rate was obtained. LDV dataset (Wei et al. 2014) is used for dataset evaluations. Kang et al. (2020) proposed a thermal

infrared image-based classification for four vehicle classes: buses, trucks, vans, and cars. 97% accuracy rate was achieved by using convolutional neural networks and SqueezeNet.

In this study, a model has been proposed for the detection of vehicle type from vehicle interior sounds using artificial intelligence methods.

### **Motivation**

Vehicle interior sound classification (VISC) can be used for many purposes. These are robotic, information security, cybercrime, traffic accident detection, vehicle type classification. Computer-aided sound classification systems are capable of extracting many features from audio signals. However, the human auditory system can detect sounds in a range of limited frequency. Computer-aided automatic sound classification systems use all values in signal frequency. Therefore, high accurate automated sound classification methods have been proposed by using computer-aided systems. Our primary motivation is to define a novel sound classification study area by using vehicles. Humans spend a long time in vehicles, and vehicles can be utilized in a crime area. Therefore, vehicle type classification is crucial for sound forensics and machine learning. Detection of vehicle type from interior car sounds It can be used for digital forensics examination. Sound files are frequently encountered in digital materials obtained from criminals. The resulting sound files are usually analyzed by manual methods. It is important to obtain information about the wanted criminal from the sound files. The method we suggest will be able to detect which type of tool the criminal is using by using sound files. It can also be used to detect noise and sound quality interior the vehicle. Interior vehicle sound classification model, which we recommend for quality detection and estimation, can be used. In order to classify vehicle types, a VISC dataset is collected, and a novel VISC model is presented in this work. Also, our second motivation is to propose a sound classification based biometric identification method for vehicles. By using this method, sound classification based vehicle sound quality systems can also be developed.

### **Our Method**

In this study, an local quintet magnitude pattern (LQMP) and INCA based VISC method are proposed for classification of the ambient sounds obtained from the cabin of eight different vehicle types used on paved roads. The study consists of four main stages. In the first stage, vehicle interior sounds are collected for different vehicle types. In the second stage, maximum pooling and the proposed LQMP method are used to generate low level, medium level, and high-level features. In the third stage, informative features are selected with INCA. In the last stage, the selected features are classified by using four conventional classifiers.

## Contributions

The main contributions are;

- This study defines a novel ESC sub-branch, and it is called as VISC. A novel hand-crafted sound classification method is presented to classify vehicles, and the presented method is tested on the collected VISC datasets. A high accurate VISC method was proposed in this work.
- A novel feature extractor is proposed in this study, and it is called as LQMP. In the local feature extractors, signum and ternary functions have generally been used for binary feature generation. In the LQMP based feature extraction method, Euclidean and City Block distances are used to generate features. The proposed LQMP is an effective feature extractor for VISC problem.

## Dataset

The used dataset is collected from the point of view (PoV) driving of different vehicle types from YouTube (Baucas and Spachos 2020). These are only vehicle interior sounds. There is no driver or any human voice. 5980 sounds were recorded with 8 classes. These vehicles were driven on asphalt road in open-air. We didn't prefer to collect interior vehicle sounds on unpaved roads in rainy weather. These sound samples were collected from 30 different vehicle videos for each class. Toplanan The file format of these data is wav. The length of the used sound samples is in the range of 3–5 seconds with 48 KHz frequency. The chosen vehicle types are bus, minibus, pickup, sport car, jeep, truck, crossover, and car (automobile). Collected dataset can be downloaded by using <https://zenodo.org/record/5606504#.YXpNrBpByUk> URL. The attributes of the collected vehicle interior sound (VIS) dataset are summarized in Table 1.

As it can be seen in Table 1, the collected dataset is a heterogeneous dataset and there are 8 classes. Therefore, this dataset is called as VISC8 dataset.

## The Proposed VISC Method

Our main goal is to propose a highly accurate and cognitive VISC method. Therefore, we proposed a multileveled VISC method. The main components of the proposed VISC method are given below.

- Feature extraction with LQMP
- Feature selection by using iterative NCA (Raghu and Sriraam 2018) (INCA).
- Classification

**Table 1.** Collected VISC dataset.

No	Class name	Number of signal	No	Class name	Number of signal
1	Bus	850	5	Jeep	600
2	Minibus	600	6	Truck	900
3	Pickup	680	7	Crossover	800
4	Sport Car	800	8	Car (C Class – 4K)	750
Total					5980

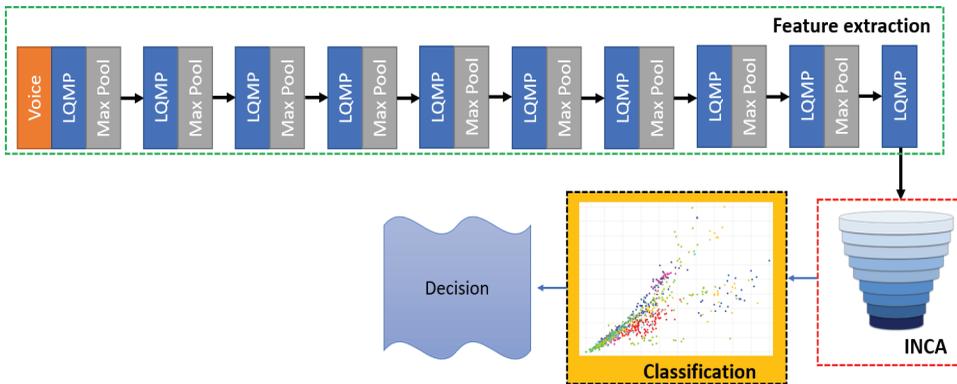
In the proposed LQMP and INCA based method, the maximum pooling operator is applied to raw vehicle interior sound 9 times. The proposed LQMP extract 128 features from raw vehicle interior sound and 9 maximum pooled signals of it. These features are fused, and 1280 features are generated totally. To select the most informative features, INCA is applied to the extracted 1280 features, and 140 most informative features are selected. kNN (Kar, Sharma, and Maitra 2015; Liao and Vemuri 2002), LD (Chen and Yang 2004), SVM (Keerthi et al. 2000), and BT (Hothorn and Lausen 2005) classifiers are used to classify these 140 most discriminative features. To better explain the proposed LQMP and INCA based VISC method, the graphical summarization of this method is shown in Figure 1 .

Moreover, pseudocode of the suggested LQMP and INCA based VISC method is denoted in Figure 2.

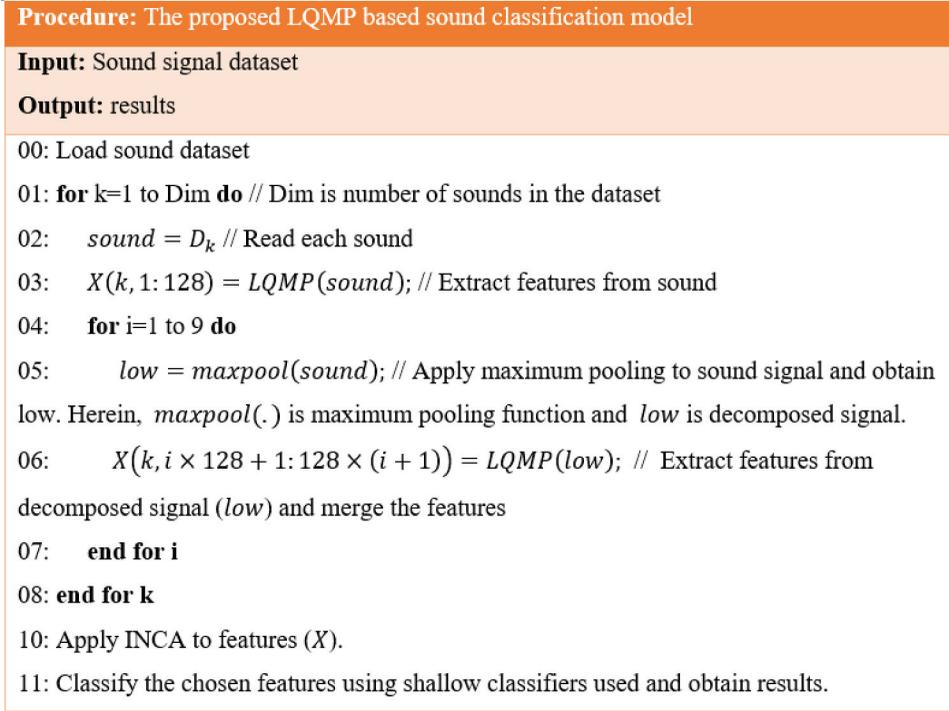
Details of the proposed VISC method are given step by step below.

**Step 1:** Apply maximum pooling (Demir et al. 2020; Jain and Bhandare 2011) to vehicle interior sound 9 times.

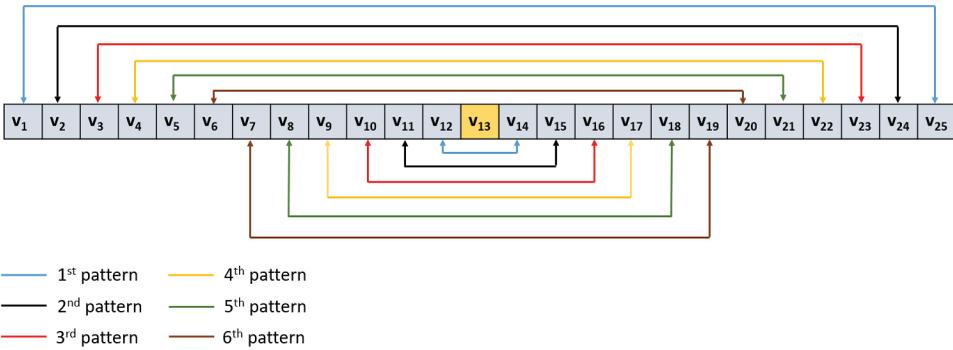
$$MPS^1(r) = \max(VIS(i), VIS(i + 1)), i = \{1, 3, \dots, L - 1\}, r = \left\{1, 2, \dots, \frac{L}{2}\right\} \tag{1}$$



**Figure 1.** Graphical summarization of the proposed LQMP and INCA based VISC method.



**Figure 2.** Pseudo code of our LQMP and INCA based VISC method.



**Figure 3.** The used patterns of the proposed LPMQ.

$$MPS^t(r) = \max(MPS^{t-1}(i), MPS^{t-1}(i+1)), t = \{2, 3, \dots, 9\} \quad (2)$$

$$\max(a, b) = \begin{cases} a, & a \geq b \\ b, & a < b \end{cases} \quad (3)$$

where *VIS* is input vehicle interior sound signal,  $MPS^t$  *t*th level maximum pooled signal, *L* represents the length of the signal, *i* and *r* define index values,

$\max(\cdot, \cdot)$  denotes maximum function and  $a, b$  are input parameters of the maximum function.

As it can be seen in Step 1, the interior voice signals are divided into 2 sized non-overlapping windows, and maximum function is applied to each window.

**Step 2:** Generate features from raw and maximum pooling sounds.

$$feat^1 = LQMP(VIS) \quad (4)$$

$$feat^h = LQMP(MPS^{h-1}), h = \{2, 3, \dots, 10\} \quad (5)$$

Step 2 defines the used feature extraction function. In this phase, LQMP is utilized as a feature generation function. The details of the proposed LQMP are given as sub-steps. The main objective of the proposed LQMP is to generate features by using 25 sized overlapping window, Euclidean, and city block distance metrics.  $5 \times 5$  sized block is used to propose this model. In this work, we used one-dimensional signals. Therefore, we used 25 sized overlapping windows. This method is a center symmetric local feature extractor. We demonstrated the recommended pattern in [Figure 3](#).

**Step 2.1:** Divide the input signal into 25 sized overlapping windows and obtain  $L - 24$  windows.

$$window = VIS(i : i + 24), i = \{1, 2, \dots, L - 24\} \quad (6)$$

**Step 2.2:** Extract Euclidean bits by using LPMQ pattern and Euclidean function.

$$ev^1(j) = \sqrt{(v(j) - v(13))^2 + (v(26 - j) - v(13))^2}, j = \{1, 2, \dots, 6\} \quad (7)$$

$$ev^2(j) = \sqrt{(v(13 - j) - v(13))^2 + (v(13 + j) - v(13))^2} \quad (8)$$

$$bit^E(j) = \begin{cases} 0, & ev^1(i) < ev^2(i) \\ 1, & ev^1(i) \geq ev^2(i) \end{cases} \quad (9)$$

where  $ev^1$  and  $ev^2$  represent first and second Euclidean values,  $v$  defines values of the non-overlapping window,  $bit^E$  denotes Euclidean bit, and  $j$  is index value.

**Step 2.3:** Generate city block bits by using Equations (10)–(12).

$$cv^1(j) = |v(j) - v(13)| + |v(26 - j) - v(13)| \quad (10)$$

$$cv^2(j) = |v(13 - j) - v(13)| + |v(13 + j) - v(13)| \quad (11)$$

$$bit^C(j) = \begin{cases} 0, & cv^1(i) < cv^2(i) \\ 1, & cv^1(i) \geq cv^2(i) \end{cases} \quad (12)$$

where  $cv^1$  and  $cv^2$  are city block values and  $bit^C$  is the city block bit.

**Step 2.4:** Calculate decimal Euclidean and city block values.

$$dev(k) = \sum_{j=1}^6 bit^E(j) * 2^{6-j}, k = \{1, 2, \dots, L - 24\} \quad (13)$$

$$dcv(k) = \sum_{j=1}^6 bit^C(j) * 2^{6-j} \quad (14)$$

where  $dev$  is decimal Euclidean value and  $dcv$  decimal city block value.

**Step 2.5:** Extract histogram of the Euclidean value and city block value. In this step, two histograms with size 64 are extracted.

$$hist^E = zeros(1, 64) \quad (15)$$

$$hist^C = zeros(1, 64) \quad (16)$$

$$hist^E(dev(k)) = hist^E(dev(k)) + 1 \quad (17)$$

$$hist^C(dcv(k)) = hist^C(dcv(k)) + 1 \quad (18)$$

where  $hist^E$  and  $hist^C$  represent Euclidean and City Block histograms. In Equations (15) and (16), histograms are defined first values of them assigned zero. Equations (17) and (18) define histogram value updating.

**Step 2.6:** Concatenate these histograms and obtain a feature vector with a size of 256.

$$feature(i) = hist^E(i), i = \{1, 2, \dots, 64\} \quad (19)$$

$$feature(i + 65) = hist^C(i) \quad (20)$$

The above steps are defined as the proposed LQMP. As it can be seen in these steps, it is a local binary pattern like feature extractor. It uses 25 sized windows center symmetrically, and kernel functions of it are Euclidean and City Block distance instead of signum or ternary function. LQMP aimed to extract informative features.

**Step 3:** Concatenate feature of each level and obtained the final feature vector with a size of 1280. Equation (21) denotes feature concatenation.

$$X((h - 1) * 128 + i) = feat^h(i), i = \{1, 2, \dots, 64\}, h = \{1, 2, \dots, 10\} \quad (21)$$

where  $X$  is the final feature vector.

**Step 4:** Select the most informative features with INCA.

$$IF = INCA(X) \quad (22)$$

In Equation (22),  $IF$  is informative features, and  $INCA$  is an iterative NCA feature selector. NCA (Raghu and Sriraam 2018) is one of the commonly preferred distance-based feature selectors and uses an optimization function (stochastic gradient descent) to calculate weights of the features. It uses a City Block distance metric based fitness function to calculate the correlation between features and target. By using this correlation value, it assigns weight to each feature. The NCA weights are positive. Therefore, it is challenging to eliminate features. The other one of the most critical problems of the NCA is to select the optimal number of features automatically. Trial and error method or parametric feature selection have been generally used in the NCA to select features. In this work, we presented an iterative feature selection model, and it is called INCA. INCA should use a classifier to calculate errors. By using minimum error, the optimal number of features is selected. The steps of the INCA are shown below.

**Step 4.1:** Normalized each column of the  $X$ .

$$X = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (23)$$

**Step 4.2:** Calculate NCA weights and index of the most informative features by using NCA function.

$$[w, index] = NCA(X) \quad (24)$$

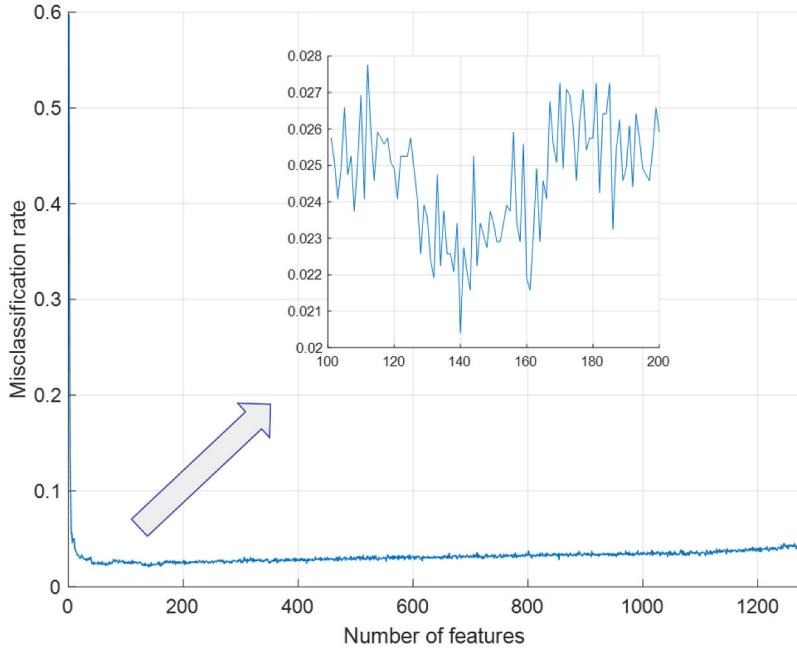
Where  $w$  represents NCA weights.

**Step 4.3:** Select features by using the index vector iteratively.

$$sf^i(i) = X(index(i)), i = \{1, 2, \dots, 1280\} \quad (25)$$

where  $sf^i$  is  $i$ th selected features.

**Step 4.4:** Calculate the error values of each selected feature by using kNN. In this step, kNN is utilized as error calculator. The parameters of the used error calculator are given as follows.  $k$  is selected as 1, and the distance metric is chosen as city



**Figure 4.** The calculated error rates by using kNN classifier.

block. kNN has a low computational cost. Therefore, we select it as an error calculator.

$$error(g) = kNN(sf^g, target) \quad (26)$$

**Step 4.5:** Select optimal features which have the minimum error.

$$[minimum, end] = min(error) \quad (27)$$

$$IF(i) = X(index(i)), i = \{1, 2, \dots, end\} \quad (28)$$

where *end* is the index value of the minimum error, we selected a range for features. Our range is 100 from 1000 features. The calculated error rates according to the number of features are shown in [Figure 4](#).

As it can be seen in [Figure 4](#), the selected final feature vector has 140 features.

**Step 5:** Use the selected most informative features as the input of the classifiers. In this step, LD, kNN, SVM, and BT classifiers are utilized as classifiers. 10-fold CV is selected to test results. The hyperparameters of the used classifiers are given as below.

*LD:* Gamma value is set as 0 (Chen and Yang 2004).

*kNN:* k is 1, and the distance metric is city block (Kar, Sharma, and Maitra 2015; Liao and Vemuri 2002).

*SVM*: A kernel function is polynomial. Polynomial order is selected 3. Kernel scale is automatic, coding is one vs. all, and constraint value (C) is chosen as 1 (Keerthi et al. 2000).

*BT*: Bag ensemble method is chosen. 30 is chosen as the number of learning cycles, the learning rate is 0.1, and 799 is selected as the maximum number of splits (Hothorn and Lausen 2005).

## Results

The experimental setup and results of the proposed LQMP and INCA based VISC method were explained in this section.

### Experimental Setup

VISC8 database was collected from YouTube PoV drive videos, and interior sounds were stored as m4a format. MATLAB2018a programming environment was used to implement our model on these sounds. The proposed LQMP and INCA based VISC method was implemented on a desktop computer, and this computer has 16 GB main memory and intel i7 7700 microprocessor. The used operating system is Windows 10.1 ultimate.

### Experimental Result

The commonly used performance evaluation metrics of the classifiers were used to determine the success of the proposed method. These metrics are classification accuracy, F1-score, and geometric mean. General mathematical notations of them were given as below.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (29)$$

$$F1 - score = \frac{2TP}{2TP + FN + FP} \quad (30)$$

$$geometric\ mean = \sqrt{\frac{TP * TN}{(TP + FN)(TN + FP)}} \quad (31)$$

where  $TP$ ,  $TN$ ,  $FP$ , and  $FN$  are the number of true positives, true negatives, false positives, and false negatives, respectively. Equations (29)–(31) are defined for two-class classification problems. However, there are 8 classes in our VISC problem. Therefore, we used Algorithm 1 to calculate these parameters for multiclass problems.

**Table 2.** The calculated performance rates (%) according to the classifier and validation techniques.

Classifier	Validation	Accuracy	F1-score	Geometric mean
LD	10-fold CV	86.64	85.69	84.95
	70:30	87.85	87.02	86.41
kNN	10-fold CV	97.88	97.67	97.61
	70:30	96.60	96.36	96.29
SVM	10-fold CV	98.38	98.23	98.21
	70:30	98.44	98.26	98.20
BT	10-fold CV	96.89	96.61	96.49
	70:30	97.10	96.89	96.88

In this section, we used two validation models and these are 10-fold cross-validation and hold-out validation. In the hold-out validation, the split ratio is chosen as 70:30. By using these validation techniques, the calculated performance metrics are tabulated in Table 2.

Table 2 denotes that the SVM classifier attained over 98% classification accuracies for both validation techniques and it is the best classifier for this problem. In Figure 5, we depicted confusion matrices of the SVM for both validatios. (a) 10-fold CV. (b) Hold-out validation and split ratio is 70:30

Procedure: Metric calculation
<b>Input:</b> Predicted values and actual outputs
<b>Output:</b> Accuracy ( <i>acc</i> ), F1-score ( <i>f1</i> ) and geometric mean ( <i>gm</i> )
01: Create confusion matrix using predicted values and actual outputs
02: $acc = 0$ ; $gm = 1$ ; // Assign the first values
03: $[m, n] = size(C)$ ; // Calculate size of the confusion matrix
04: <b>for</b> $k=1$ to $m$ <b>do</b>
05: $r(k) = \frac{C(k,k)}{sum(C(k,:))}$ // Calculate class-wise recall
06: $p(k) = \frac{C(k,k)}{sum(C(:,k))}$ // Calculate class-wise precision
07: $gm = gm \times r(k)$
08: $acc = acc + C(k, k)$ // Calculate accuracy mean
09: <b>end for</b> $k$
10: $Recall = mean(r)$ ; // Calculate unweighted average recall
11: $Precision = mean(p)$ ; // Calculate average precision
12: $Accuracy = acc/sum(C)$ ; // Calculate average precision
13: $F1 = 2 \times Recall \times Precision / (Recall + Precision)$ ; // Calculate F1-score
14: $gm = \sqrt[m]{gm}$ // Calculate geometric mean

**Figure 5.** The procedure of metric calculation.

Predicted Class

		1	2	3	4	5	6	7	8
Output Class	1	842 99.06%	0 0%	0 0%	0 0%	0 0%	1 0.12%	6 0.70%	1 0.12%
	2	0 0%	595 99.17%	3 0.50%	0 0%	1 0.17%	0 0%	0 0%	1 0.17%
	3	1 0.15%	1 0.15%	665 97.79%	1 0.15%	5 0.74%	0 0%	1 0.15%	6 0.88%
	4	2 0.25%	0 0%	0 0%	792 99%	4 0.50%	0 0%	0 0%	2 0.25%
	5	2 0.33%	17 2.83%	5 0.83%	0 0%	563 93.83%	1 0.17%	5 0.83%	7 1.17%
	6	1 0.11%	2 0.22%	0 0%	0 0%	0 0%	897 99.67%	0 0%	0 0%
	7	0 0%	0 0%	4 0.50%	0 0%	1 0.13%	0 0%	791 98.87%	4 0.50%
	8	0 0%	0 0%	0 0%	0 0%	5 0.67%	0 0%	7 0.93%	738 98.40%

(a) 10-fold CV

Predicted Class

		1	2	3	4	5	6	7	8
Output Class	1	253 99.22%	1 0.39%	0 0%	0 0%	0 0%	1 0.39%	0 0%	0 0%
	2	0 0%	180 100%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%
	3	1 0.49%	0 0%	199 97.55%	0 0%	2 0.98%	0 0%	0 0%	2 0.98%
	4	0 0%	0 0%	0 0%	239 99.58%	0 0%	0 0%	0 0%	1 0.42%
	5	0 0%	8 4.44%	2 1.11%	0 0%	165 91.67%	1 0.56%	3 1.67%	1 0.56%
	6	0 0%	0 0%	0 0%	0 0%	0 0%	270 100%	0 0%	0 0%
	7	0 0%	0 0%	3 1.25%	0 0%	0 0%	0 0%	235 97.92%	2 0.83%
	8	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	225 100%

(b) Hold-out validation and split ratio is 70:30

**Figure 6.** Confusion matrices of the proposed LQMP and INCA based VISC method using SVM classifier with (a) 10-fold CV, (b) hold-out validation.

Figure 6 denotes the calculated confusion matrices using SVM classifier. Per Figure 5, the best accurate class is 6th class using 10-fold CV (see Figure 5) since it achieved 99.67% classification accuracy. By testing the proposed LQMP and INCA based VISC method with 70:30 split ratio, our model reached 100% (perfect accuracy) using 2nd, 6th, and 8th classes.

## Discussions

In this work, we presented a novel sub-study area of the ESC, and it is called as VISC. VISC is the crucial study area because it can be used in both ESC and biometric identification of the vehicles. We presented the first study of this area. Therefore, we collected a dataset from YouTube PoV drive videos. Sounds of these videos were recorded. We presented a hand-crafted feature extraction model, and this model has multilevel. We inspired deep networks for instance AlexNet (Krizhevsky, Sutskever, and Hinton 2012), GoogLeNet (Szegedy et al. 2015), ResNet (He et al. 2016). A new feature generation function was presented to achieve high accuracy. The proposed feature generation function is LQMP. The main point of the proposed LQMP is to extract informative features by using city block and Euclidean distance based binary feature-extraction functions, and it generates 128 features. The computational cost of the proposed LQMP and maximum pooling based feature generation method was calculated as  $O(n \log n)$  because maximum pooling operator decomposed signal in each level and time cost of the LQMP is directly depend on the length of signal ( $O(n)$ ). An optimal number of features were selected by using INCA feature selector. The proposed LQMP and maximum pooling based feature generation method extract 1280 features and INCA select 140 most informative features of these 1280 features. kNN was utilized as an error value calculator, and we determined the range for the number of features. Because this range determines the number of executions for kNN. kNN also has low computational complexity ( $O(nd)$ ). Therefore, it was used in INCA as a loss calculator. Four conventional classifiers were used to denote the efficiency of these 140 most informative features. 86.64%, 97.88%, 98.38%, and 96.89% classification accuracies were calculated by using LD, BT, kNN, and SVM classifiers with 10-fold cross-validation. Moreover, the suggested model have been tested with hold-out validation and our proposal achieved 87.85%, 96.60%, 98.44%, and 97.10% classification accuracies were calculated by using LD, BT, kNN, and SVM classifiers consecutively. Benefits of the proposed LQMP and INCA based VISC method are;

- VISC is a new study area, and this study area effected automotive, machine learning, and crime investigation.
- A novel feature extraction method (LQMP and maximum pooling) was presented.

- An automated feature selector (INCA) was presented to overcome the optimal feature selection problem of the NCA.
- By using these methods (LQMP based feature extraction and INCA feature selection), a high accurate VISC method was presented, and this method reached 98.44% classification accuracy for the collected 800 sounds of the VISC8 dataset.
- The proposed LQMP and INCA based method is a robust classification method because the results were obtained by using 10-fold cross-validation.
- Results of four classifiers were presented to evaluate the proposed LQMP and INCA based VISC method comprehensively.

## Conclusions

We propose a novel sound classification study area in this paper. Our main aim is to classify the type of vehicles with high classification accuracy. Therefore, a novel dataset was collected, and it is called as VISC8 dataset. A new hand-crafted feature generation and iterative feature selection based method is proposed to test this dataset. Maximum pooling is used to create levels, and LQMP extracts 128 features from each level. 140 most informative features were selected with INCA, and these features were forwarded to LD, kNN, SVM, and BT classifiers. The best accuracy rates of the whole classifiers were higher than 86%, and 98.44% classification rate was achieved using SVM with 70:30 split ratio. Our findings clearly illustrated that the vehicles could be classified by using their interior sounds.

In the future works, our method can be motivated by other researchers, and novel big and parametric datasets can be collected. In the sound collection phase, several parameters such as weather, road type can be used, and novel learning methods can be presented to classify these datasets. Also, the proposed LQMP and INCA based methods can be used to create learning methods.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

Erhan Akbal  <http://orcid.org/0000-0002-5257-7560>

Turker Tuncer  <http://orcid.org/0000-0002-1425-4664>

Sengul Dogan  <http://orcid.org/0000-0001-9677-5684>

## References

ACEA Report Vehicles in Use Europe 2019. 2019.

- Akhtar, S., M. Elshafei-Abmed, and M. S. Ahmed. 2001. Detection of helicopters using neural nets. *IEEE Transactions on Instrumentation and Measurement* 50 (3):749–56. doi:10.1109/19.930449.
- Al-Ali, A. K. H., B. Senadji, and G. R. Naik. 2017. Enhanced forensic speaker verification using multi-run ICA in the presence of environmental noise and reverberation conditions. Paper Presented at the 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA) 12-14 September 2017 Kuching, Malaysia.
- Baucas, M. J., and P. Spachos. 2020. Using cloud and fog computing for large scale IoT-based urban sound classification. *Simulation Modelling Practice and Theory* 101:102013. doi:10.1016/j.simpat.2019.102013.
- Boucher, C., and J.-C. Noyer. 2012. Automatic detection of topological changes for digital road map updating. *IEEE Transactions on Instrumentation and Measurement* 61 (11):3094–102. doi:10.1109/TIM.2012.2203873.
- Chang, P.-C., Y.-S. Chen, and C.-H. Lee. 2021. MS-SincResnet: Joint learning of 1D and 2D kernels using multi-scale SincNet and ResNet for music genre classification. ICMR '21: Proceedings of the 2021 International Conference on Multimedia Retrieval 21-24 August 2021 Taipei Taiwan.
- Chen, S., and X. Yang. 2004. Alternative linear discriminant classifier. *Pattern Recognition* 37 (7):1545–47. doi:10.1016/j.patcog.2003.11.008.
- Demir, S., S. Key, T. Tuncer, and S. Dogan. 2020. An exemplar pyramid feature extraction based humerus fracture classification method. *Medical Hypotheses* 109663. doi:10.1016/j.mehy.2020.109663.
- Duarte, M. F., and Y. H. Hu. 2004. Vehicle classification in distributed sensor networks. *Journal of Parallel and Distributed Computing* 64 (7):826–38. doi:10.1016/j.jpdc.2004.03.020.
- Gramatikov, B. I., S. Rangarajan, K. Irsch, and D. L. Guyton. 2016. Attention attraction in an ophthalmic diagnostic device using sound-modulated fixation targets. *Medical Engineering & Physics* 38 (8):818–21. doi:10.1016/j.medengphy.2016.05.004.
- Hashimoto, T., and H. Takao. 1990. Subjective estimation of running car interior noise, part 2: The relation between subjective and physical values of noise. *Transactions of the Automotive Engineers of Japan* (43):129–33.
- He, K., X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. Paper Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 27-30 June 2016 Las Vegas, NV, USA.
- Hothorn, T., and B. Lausen. 2005. Bundling classifiers by bagging trees. *Computational Statistics & Data Analysis* 49 (4):1068–78. doi:10.1016/j.csda.2004.06.019.
- Huang, J., C. Chen, G. Guo, Z. Zhang, and Z. Li. 2019. A calculation model for typical data center cooling system. *Journal of Physics: Conference Series*, Volume 1304, The 2nd International Conference on Energy, Electrical and Power Engineering 25–28 June 2019 Berkley, USA.
- Jain, Y. K., and S. K. Bhandare. 2011. Min max normalization based data perturbation method for privacy protection. *International Journal of Computer & Communication Technology* 2 (8):45–50.
- Jeon, J. Y., J. You, C. I. Jeong, S. Y. Kim, and M. J. Jho. 2011. Varying the spectral envelope of air-conditioning sounds to enhance indoor acoustic comfort. *Building and Environment* 46 (3):739–46. doi:10.1016/j.buildenv.2010.10.005.
- Kang, Q., H. Zhao, D. Yang, H. S. Ahmed, and J. Ma. 2020. Lightweight convolutional neural network for vehicle recognition in thermal infrared images. *Infrared Physics & Technology* 104:103120. doi:10.1016/j.infrared.2019.103120.
- Kar, S., K. D. Sharma, and M. Maitra. 2015. Gene selection from microarray gene expression data for classification of cancer subgroups employing PSO and adaptive K-nearest neighborhood technique. *Expert Systems with Applications* 42 (1):612–27. doi:10.1016/j.eswa.2014.08.014.

- Keerthi, S. S., S. K. Shevade, C. Bhattacharyya, and K. R. Murthy. 2000. A fast iterative nearest point algorithm for support vector machine classifier design. *IEEE Transactions on Neural Networks* 11 (1):124–36. doi:10.1109/72.822516.
- Khan, M. K., M. Zakariah, H. Malik, and K.-K. R. Choo. 2018. A novel audio forensic data-set for digital multimedia forensics. *The Australian Journal of Forensic Sciences* 50 (5):525–42. doi:10.1080/00450618.2017.1296186.
- Ko, K., J. Park, D. K. Han, and H. Ko. 2019. Channel and frequency attention module for diverse animal sound classification. *IEICE Transactions on Information and Systems* 102 (12):2615–18. doi:10.1587/transinf.2019EDL8128.
- Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 December 2012 Lake Tahoe Nevada.
- Lemaitre, G., and P. Sussini. 2019. Timbre, sound quality, and sound design Kai Siedenburg, Charalampos Saitis, Stephen McAdams, Arthur N. Popper, Richard R. Fay . In *Timbre: Acoustics, perception, and cognition*, 245–72 978-3-030-14831-7 doi:10.1007/978-3-030-14832-4 . Springer Nature Switzerland AG 2019: Springer.
- Liao, Y., and V. R. Vemuri. 2002. Use of k-nearest neighbor classifier for intrusion detection. *Computers & Security* 21 (5):439–48. doi:10.1016/S0167-4048(02)00514-X.
- Liu, C., L. Feng, G. Liu, H. Wang, and S. Liu. 2021. Bottom-up broadcast neural network for music genre classification. *Multimedia Tools and Applications* 80 (5):7313–31. doi:10.1007/s11042-020-09643-6.
- Lurz, F., S. Lindner, S. Linz, S. Mann, R. Weigel, and A. Koelpin. 2017. High-speed resonant surface acoustic wave instrumentation based on instantaneous frequency measurement. *IEEE Transactions on Instrumentation and Measurement* 66 (5):974–84. doi:10.1109/TIM.2016.2642618.
- Manders, E.-J., G. Biswas, P. J. Mosterman, L. A. Barford, and R. J. Barnett. 2000. Signal interpretation for monitoring and diagnosis, a cooling system testbed. *IEEE Transactions on Instrumentation and Measurement* 49 (3):503–08. doi:10.1109/19.850384.
- Mayer, P., M. Magno, and L. Benini. 2019. Self-sustaining acoustic sensor with programmable pattern recognition for underwater monitoring. *IEEE Transactions on Instrumentation and Measurement* 68 (7):2346–55. doi:10.1109/TIM.2018.2890187.
- Mazarakis, G. P., and J. N. Avaritsiotis. 2007. Vehicle classification in sensor networks using time-domain signal processing and neural networks. *Microprocessors and Microsystems* 31 (6):381–92. doi:10.1016/j.micpro.2007.02.005.
- Mishra, M., H. Menon, and A. Mukherjee. 2018. Characterization of \$ S\_1 \$ and \$ S\_2 \$ heart sounds using stacked autoencoder and convolutional neural network. *IEEE Transactions on Instrumentation and Measurement* 68 (9):3211–20. doi:10.1109/TIM.2018.2872387.
- More, S., and P. Davies. 2010. Human responses to the tonalness of aircraft noise. *Noise Control Engineering Journal* 58 (4):420–40. doi:10.3397/1.3475528.
- Nanni, L., Y. M. Costa, D. R. Lucio, C. N. Silla, Jr, and S. Brahnam. 2017. Combining visual and acoustic features for audio classification tasks. *Pattern Recognition Letters* 88:49–56. doi:10.1016/j.patrec.2017.01.013.
- Patidar, S., and R. B. Pachori. 2014. Classification of cardiac sound signals using constrained tunable-Q wavelet transform. *Expert Systems with Applications* 41 (16):7161–70. doi:10.1016/j.eswa.2014.05.052.
- Raghu, S., and N. Srimaam. 2018. Classification of focal and non-focal EEG signals using neighborhood component analysis and machine learning algorithms. *Expert Systems with Applications* 113:18–32. doi:10.1016/j.eswa.2018.06.031.
- Randhawa, S. K., and M. Singh. 2015. Classification of heart sound signals using multi-modal features. *Procedia Computer Science* 58:165–71. doi:10.1016/j.procs.2015.08.045.

- Salamon, J., C. Jacoby, and J. P. Bello. 2014. A dataset and taxonomy for urban sound research. MM '14: Proceedings of the 22nd ACM international conference on Multimedia 3-7 November 2014 New York, United States.
- Scanlon, P., D. F. Kavanagh, and F. M. Boland. 2012. Residual life prediction of rotating machines using acoustic noise signals. *IEEE Transactions on Instrumentation and Measurement* 62 (1):95–108. doi:10.1109/TIM.2012.2212508.
- Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, and A. Rabinovich. 2015. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 15 October 2015 Boston, MA.
- Takada, M., H. Mori, S. Sakamoto, and S.-I. Iwamiya. 2019. Structural analysis of the value evaluation of vehicle door-closing sounds. *Applied Acoustics* 156:306–18. doi:10.1016/j.apacoust.2019.07.025.
- Västfjäll, D., M.-A. Gulbol, M. Kleiner, and T. Gärling. 2002. Affective evaluations of and reactions to exterior and interior vehicle auditory quality. *Journal of Sound and Vibration* 255 (3):501–18. doi:10.1006/jsvi.2001.4166.
- Wang, Y., G. Shen, H. Guo, X. Tang, and T. Hamade. 2013. Roughness modelling based on human auditory perception for sound quality evaluation of vehicle interior noise. *Journal of Sound and Vibration* 332 (16):3893–904. doi:10.1016/j.jsv.2013.02.030.
- Wei, J., C.-H. Liu, Z. Zhu, L. R. Cain, and V. J. Velten. 2019. Vehicle engine classification using normalized tone-pitch indexing and neural computing on short remote vibration sensing data. *Expert Systems with Applications* 115:276–86. doi:10.1016/j.eswa.2018.07.073.
- Wei, J., N. Li, X. Xia, X. Chen, F. Peng, G. E. Besner, and J. Feng. 2014. Effects of lipopolysaccharide-induced inflammation on the interstitial cells of cajal. *Cell and Tissue Research* 356 (1):29–37. doi:10.1007/s00441-013-1775-7.
- Yang, L., and H. Zhao. 2021. Sound classification based on multihead attention and support vector machine. *Mathematical Problems in Engineering* 2021.