Fırat Üniversitesi Deneysel ve Hesaplamalı Mühendislik Dergisi

Ses Analizi Yoluyla Doğru Ev İçi Konumu Sınıflandırması: 1D-ILQP Yaklaşımı

Nura ABDULLAHI[1], Erhan AKBAL[2*], Sengul DOGAN[3], Turker TUNCER[4], Umut ERMAN[5]
[1,2,3,4,5]Department of Digital Forensics Engineering, Technology Faculty, Firat University, Elazig, Türkiye.
[1]191144116@firat.edu.tr, [2]erhanakbal@firat.edu.tr, [3]sdogan@firat.edu.tr, [4]turkertuncer@firat.edu.tr,
[5]221144111@firat.edu.tr

Geliş Tarihi: 19.01.2024
Kabul Tarihi: 02.05.2024
Düzeltme Tarihi: 17.03.2024
doi: https://doi.org/10.62520/fujece.1422119
Araştırma Makalesi

Alıntı: N. Abdullahi, E. Akbal, Ş. Dogan, T. Tuncer ve U. Erman, "Ses analizi yoluyla doğru ev içi konumu sınıflandırması: 1D-ILQP yaklaşımı", Fırat Üni. Deny. ve Hes. Müh. Derg., vol. 4, no 1, pp. 12-29, Şubat 2025.

Öz

Ev ortamlarındaki insan faaliyetlerinin tespit edilmesi, makine öğrenimi alanında temel bir zorluk teşkil etmektedir. Geleneksel olarak sensörler ve video kameralar, insan faaliyetinin tespitinde birincil araçlar olarak hizmet vermiştir. Ancak çalışmamız, çevresel ses sinyallerinin analizi yoluyla ev içi konumlarını belirlemeye yönelik yenilikçi hedefe sahiptir. Sonuç olarak, sekiz farklı lokasyondan gelen verileri kapsayan kapsamlı bir ses veri seti toplanmıştır. Bu ses veri kümesini kullanarak otomatik ev konumu algılamayı etkinleştirmek için, hassasiyete ve minimum hesaplama yüküne odaklanarak hafif bir makine öğrenimi modeli kullanılmıştır. Yaklaşımımızın temelinde, tek boyutlu Geliştirilmiş Yerel Dörtlü Model (GYDM) olarak adlandırılan yerel bir özellik oluşturucunun tanıtılması yer almaktadır. Bu yöntem, akustik sinyallerden dokusal özellikler üreterek özellik çıkarma sürecinde merkezi bir rol oynar. Yüksek seviyeli dokusal özelliklerin çıkarılmasını kolaylaştırmak için, sinyalleri ayrıştırmak için maksimum havuzlama uygulayarak evrişimli sinir ağı mimarisini taklit edilmiştir. Önerilen GYDM, orijinal sinyalin yanı sıra her ayrıştırılmış frekans bandından dokusal özelliklerini çıkarmaktadır. Daha sonra Komşu Bileşen Analizi tekniğini kullanarak en iyi 100 özellik seçilmiştir.Modelimizin son adımı sınıflandırmayı içermektedir. Bu aşamada karar ağaçları, doğrusal diskriminant analizi, ikinci dereceden diskriminant analizi, Naive Bayes, destek vektör makineleri, k-en yakın komşu, torbalanmış ağaçlar ve yapay sinir ağları dahil olmak üzere bir dizi sınıflandırıcı kullanılmıştır. Sonuçlar kapsamlı bir değerlendirmeye tabi tutulmuş ve tüm sınıflandırıcılar %80'in üzerinde sınıflandırma doğruluğuna ulaşmıştır. Özellikle k-en yakın komşu sınıflandırıcı, %99,75 gibi etkileyici bir değere ulaşarak en yüksek sınıflandırma doğruluğu sağlamıştır. Bulgularımız, GYDM'ye dayanan önerilen ses sınıflandırma modelinin, ev konumu ses veri setine uygulandığında oldukça tatmin edici sonuçlar verdiğini açıkça göstermektedir.

Anahtar kelimeler: Ev konum tespiti, Komşu bileşen analizi, Ses sınıflandırma, Makine öğrenmesi

*Yazışılan yazar

İntihal Kontrol: Evet – Turnitin
Şikayet: fujece@firat.edu.tr

Firat Univ Jour. of Exp. and Comp. Eng., 4(1), 12-29, 2025

Firat University Journal of Experimental and Computational Engineering

Accurate Indoor Home Location Classification through Sound Analysis: The 1D-ILQP Approach

Nura ABDULLAHI[1], Erhan AKBAL[*2], Sengul DOGAN[3], Turker TUNCER[4], Umut ERMAN[5]

[1,2,3,4,5] Department of Digital Forensics Engineering, Technology Faculty, Firat University, Elazig, Turkey.

[1]191144116@firat.edu.tr, [2]erhanakbal@firat.edu.tr, [3]sdogan@firat.edu.tr, [4]turkertuncer@firat.edu.tr, [5]221144111@firat.edu.tr

Received: 19.01.2024
Accepted: 02.05.2024

Revision: 17.03.2024

doi: https://doi.org/10.62520/fujece.1422119
Research Article

Citation: N. Abdullahi, E. Akbal, Ş. Dogan, T. Tuncer and U. Erman, "Accurate ındoor home location classification through sound analysis: the 1D-ILQP approach", Firat Univ. Jour.of Exper. and Comp. Eng., vol. 4, no 1, pp. 12-29, February 2025.

Abstract

Detecting human activities within domestic environments constitutes a fundamental challenge in machine learning. Conventionally, sensors and video cameras served as primary tools for human activity detection. However, our work is oriented towards the innovative objective of ascertaining home locations by analyzing environmental sound signals. Consequently, we compiled a comprehensive sound dataset from eight distinct locations. To enable automatic home location detection using this sound dataset, we employed a lightweight machine learning model designed with a paramount focus on precision and minimal computational overhead. At the core of our approach is the introduction of a local feature generator, referred to as the one-dimensional Improved Local Quadruple Pattern (1D-ILQP). This novel 1D-ILQP plays a central role in the feature extraction process, generating textural features from the acoustic signals. To facilitate the extraction of high-level textural features, we emulated the convolutional neural network (CNN) architecture, applying maximum pooling to decompose signals. The suggested 1D-ILQP extracts textural features from each decomposed frequency band as well as the original signal. Subsequently, we selected the top 100 features using the Neighborhood Component Analysis (NCA) technique. The final step of our model involves classification, wherein we employed a range of classifiers, including decision trees, linear discriminant analysis, quadratic discriminant analysis, Naive Bayes, support vector machines, k-nearest neighbor, bagged trees, and artificial neural networks. We subjected the results to a comprehensive evaluation, and all classifiers achieved classification accuracies exceeding 80%. Notably, the k-nearest neighbor classifier delivered the highest classification accuracy, reaching an impressive 99.75%. Our findings unequivocally demonstrate that the proposed sound classification model, based on the 1D-ILQP, has yielded highly satisfactory results when applied to the home location sound dataset.

Keywords: Home location detection, 1D-ILQP, Neighborhood component analysis, Sound classification, Machine learning

[*]Corresponding author

Plagiarism Checks: Yes – Turnitin
Complaints: fujece@firat.edu.tr

1. Introduction

Analyzing human behavior and activities holds significant potential across various domains [1]. Behavioral data can be invaluable for product development, design, and understanding human behavior in diverse settings. Humans spend the majority of their lives in a variety of locations, such as homes, shopping malls, offices, and restaurants. Analyzing usage patterns of these places can shed light on the complex relationships between individuals and their environments [2, 3]. Traditionally, cameras and sensors have been the preferred tools for monitoring human activities. However, these methods are often costly and present challenges in data analysis. Several studies have employed image data and intelligent techniques to describe human activities, but such approaches are susceptible to environmental conditions, such as variations in lighting, physical obstructions, and camera blind spots [4-8].

Numerous studies explored the detection of human activities using sensor data [1, 9-13]. Wearable technologies and mobile device sensors gained widespread use in this context [14-17]. However, a primary limitation of sensor-based methods is their dependency on the presence of sensors on individuals, which is not always feasible, especially within the confines of one's home [18]. When people are at home, they typically do not carry mobile devices or wear wearable technology products, making these methods less effective for indoor human activity and location data collection [19].

Sound-based detection of human activities and locations offers a more cost-effective and practical alternative to video recording and sensor technologies. Human behavior can be discerned from the sounds they produce or the sounds of objects they interact with, as these sounds often exhibit location-specific patterns [20]. Sound is readily propagated and can be captured at a distance, making it a viable option for monitoring human activities without the need for individuals to wear specific sensor-equipped devices [19].

Intelligent sound recognition systems employ algorithms to learn and interpret sound patterns, extracting valuable information from sound signals [21]. Sound recognition methods are typically categorized based on the type of sound signal and their intended purpose. Common classes include environmental sound determination [22-27], music recognition [28, 29], speaker recognition [30, 31], and emotional state detection [32-34]. Additionally, acoustic event detection focuses on identifying specific events within the acoustic environment, such as falls, breaks, and impacts [35-37], while acoustic scene classification aims to classify environments, events, or behaviors using multiple concurrent sounds in the environment [38-40]. This classification process helps differentiate environments such as crime scenes, schools, restaurants, barbershops, and cafes [41, 42].

Indoor activity and location detection intersect with both event detection and acoustic scene classification. Human indoor activities generate sound signals with varying frequencies and decibels. For example, sounds produced during activities like cooking or cleaning in a home may exhibit characteristics of both acoustic event recognition and acoustic scene classification. Nonetheless, indoor activities possess unique sound signal features specific to their respective locations [20].

In the existing literature, there are limited studies that employ sound signals for the classification of human activities and location recognition/classification. For instance, Jung and Chi [43] introduced a method for classifying human activities, utilizing a dataset of 10 different human activity sounds collected from YouTube. They extracted features from the sound signals using the Log Mel filter bank and achieved an 87.6% accuracy rate using a residual convolutional neural network. Galván-Tejada et al. [44] proposed a method to recognize the indoor location of humans based on sound signals generated during activities within a house. They used 11 different human activity sounds recorded in four distinct rooms of a house and achieved a classification accuracy of 95% using the random forest model. Do et al. [45] presented a sound-based human activity monitoring model by recognizing sound events in a home environment, achieving a classification accuracy of 92.41% for 12 activities across six locations using a two-level dynamic Bayesian network. Wang et al. [46] conducted a comparative study on different feature extraction and machine learning techniques for indoor environmental noise classification. They used 2500 indoor audio signals across five classes and demonstrated a 78%

classification accuracy using the LPCC and SVM model. Mesaros et al. [47] developed a system for acoustic event detection in real-life environments using hidden Markov models and achieved 24% accuracy in classifying sound events belonging to 61 isolated classes.

1.1. Motivation and our work

Sound activity classification is a critical aspect of building and design construction. Extracting meaningful information from environmental sounds is a valuable pursuit in this field. Deep learning models and networks are frequently employed to achieve high classification accuracy in sound categorization. However, these models are computationally expensive. Therefore, the development of lightweight learning models has emerged as a significant research challenge.

The primary objective of this research is to introduce a lightweight sound classification model. To achieve this, a novel hand-crafted feature extraction function has been devised to capture highly distinctive features. However, this hand-crafted extraction function is limited in its ability to generate high-level features. To address this limitation, we have emulated a deep learning network, specifically a convolutional neural network, to create hierarchical representations. These hierarchical representations enable the extraction of high-level features. For an accurate model, it is crucial to employ an effective feature selection function. We have applied Neighborhood Component Analysis (NCA) [48] to select the most distinctive features from the extracted set.

Additionally, we collected a new dataset specific to home activities, comprising eight distinct classes of sounds. Our proposed lightweight learning model has been applied to this newly collected sound dataset. To establish benchmark results, we have employed eight shallow classifiers for the classification process.

1.2. Novelties and contributions

In this research, we have introduced a novel Environmental Sound Classification (ESC) model utilizing environmental sounds from domestic settings. The key contributions of our study include:

- We collected a new and comprehensive sound dataset specifically designed for detecting home locations. This dataset serves as a valuable resource for the development and evaluation of our classification model.
- We have introduced a novel feature extraction function named 1D-ILQP. This function is instrumental in extracting informative features from the sound data, enabling the accurate classification of environmental sounds.
- Building upon the 1D-ILQP feature extraction, we have developed a new classification model. This model leverages the capabilities of 1D-ILQP and has demonstrated remarkable performance, achieving a classification accuracy exceeding 99%. Notably, this high accuracy is accomplished using only two shallow classifiers, highlighting the efficiency and effectiveness of our approach.

These contributions collectively enhance the field of environmental sound classification, particularly in the context of home locations, and offer valuable insights for future research in this area.

1.3. Organization

The organization of the remainder of this research is presented as follows: Section 2 details the dataset, Section 3 explains the proposed one-dimensional improved local quadruple pattern, and Section 4 presents the proposed feature engineering model. The results of this feature engineering model are given in Section 5. These results are discussed in Section 6, while the limitations and future work are presented in Section 7. Finally, Section 8 concludes the proposed research.

2. Dataset

In order to evaluate the performance of the proposed method, a new dataset containing indoor sounds was meticulously assembled. Different indoor locations within a home give rise to a diverse array of sounds produced by human activities. For instance, the kitchen might feature sounds associated with cooking, eating, and washing dishes, while the bedroom may exhibit sounds related to sleep. Each of these produced sound signals possesses distinct characteristics, which render them amenable to classification based on their unique features.

In our study, we identified eight distinct indoor locations within the home. For each location, we further delineated various sub-activities and selected recordings of these sub-activities. To accomplish this, we sourced material from open-access online video and audio platforms, including YouTube. Importantly, the audio and video recordings used in our dataset originate from a variety of recording devices and media sources. This diversity is intentional, as it reflects real-world conditions where data may be collected from different environments and devices. Collecting data from diverse sources is essential to ensure that our method's performance is robust and can generalize effectively across various settings.
For our dataset, we incorporated 50 different videos for each of the selected locations. This process yielded a total of 500 sample sound signals for each location. Consequently, we created a balanced dataset, which comprises a total of 4,000 sound samples distributed across eight distinct location classes. These classes are as follows: (1) Bathroom, (2) Bedroom, (3) Dining room, (4) Dressing room, (5) Kitchen, (6) Living room, (7) Toilet, and (8) Washing room. Each audio sample is typically in the range of 1 to 2 seconds in duration, and all recordings have a signal frequency of 48 kHz.
A detailed breakdown of the dataset's contents is presented in Table 1, providing a comprehensive overview of the dataset's composition and structure.

Table 1. Detail of collected sound dataset

Class No	Location Name	Class Activity Content	Number of used Recording	Number of sample sound signal
1	Bathroom	Showering, brushing teeth, hair care	50	500
2	Bedroom	Sleeping, breathing, and snore sounds	50	500
3	Dining room	Eating, drinking, and other eating sounds (spoon and fork)	50	500
4	Dressing room	Folding clothes, measuring, and hand movements	50	500
5	Kitchen	Cleaning and water flushing	50	500
6	Living room	Singing, Talking, Studying, Music, Clicking sound of keyboard and mouse	50	500
7	Toilet	Flushing	50	500
8	Washing room	Washing Machine sounds (watering, draining and spinning), Ironing	50	500
		TOTAL	400	4.000

3. The Proposed One-Dimensional Improved Local Quadruple Pattern

In our research, we have introduced a novel feature extraction function named 1D-ILQP. The primary purpose of this function is to generate discriminative features from a given signal. The 1D-ILQP function operates by employing an overlapping block with a length of 16 and applying the signum function to create features. As a result, this feature extractor produces three map signals, each represented with eight bits. The cumulative outcome of this histogram-based extraction process amounts to 768 distinct features.

To elucidate the operation of the 1D-ILQP feature extraction function, we outline its steps below for a clearer understanding:

Step 1: Generate overlapping blocks with overlapping blocks with a length of 16.
Step 2: Apply vector to matrix transformation to obtain a 4 × 4 sized matrixes. The prime objective of this step is to apply the proposed pattern. Schematic description of the presented 1D-ILQP is shown in Figure 1.

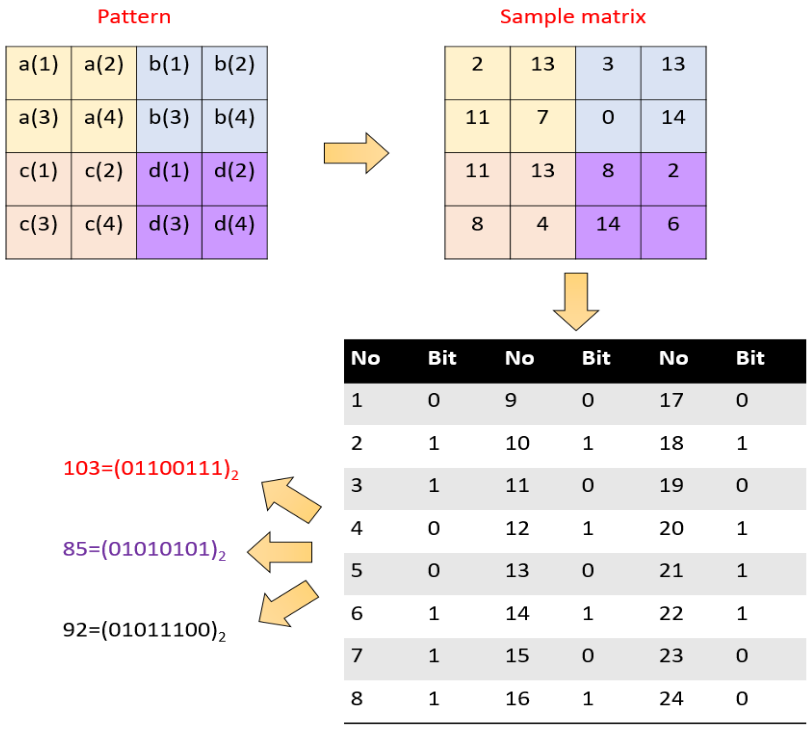

Figure 1. Binary feature extraction and decimal values creation diagram of the proposed 1D-ILQP

The proposed 1D-ILQP feature extraction function employs a binary feature extraction process, which subsequently leads to the creation of decimal values. The following diagram illustrates the binary feature extraction and decimal values creation process:

Here's a step-by-step description of this process:

- A 4x4-sized matrix is divided into four parts, each with a size of 2x2. These parts are denoted as 'a,' 'b,' 'c,' and 'd.'
- By comparing these four parts ('a,' 'b,' 'c,' and 'd'), 24 bits are extracted. These 24 bits are used to form three distinct groups of bits.
- These three bit groups serve as the basis for creating three map signals.
- Histograms of these map signals are generated to produce a feature vector.
- The three histograms are then combined, resulting in a feature vector with a total length of 768.

This feature extraction method yields a comprehensive set of features that can be utilized for subsequent analysis and classification tasks.
As can be seen Figure 1, the used pattern of the 1D-ILQP has been summarized.

Step 3: Extract binary features by using the quadruple pattern (see Figure 2). Pseudocode of the bit extraction process has been given in Algorithm 1.
Algorithm 1. Bit (binary features) generation of the proposed 1D-ILQP

Input: Matrix (m) **Output:** Bit vector (bit)
Assign a, b, c and d values using m. **for** j=1 to 4 **do** $bit(j) = a(j) \geq b(j);$ $bit(j+4) = a(j) \geq c(j);$ $bit(j+8) = a(j) \geq d(j);$ $bit(j+12) = b(j) \geq c(j);$ $bit(j+16) = b(j) \geq d(j);$ $bit(j+20) = c(j) \geq d(j);$ **end for j**

Step 4: Create three big groups with a length of eight like local binary pattern. These creation process is denoted in Equation 1.

$$b^g = bit\big(j + 8 \times (g-1)\big), j \in \{1,2,\ldots,8\}, g \in \{1,2,3\} \quad (1)$$

Herein, b^g defines g^th^ bit group.

Step 5: Generate three feature map signals using the generated bit groups.

$$map^g(i) = \sum_{j=1}^{8} b^g(j) \times 2^{j-1} \quad (2)$$

where map^g represents g^{th} feature map signal.

Step 6: Extract histogram of each map signal.

Step 7: Merge the extracted histogram to create feature vector with a length of 768.

$$fv = H^g\big(h + 256 \times (g-1)\big), h \in \{1,2,\ldots,256\} \quad (3)$$

Herein H^g is histogram of the g^{th} map signal and fv defines feature vector with a length of 768.

4. Our Proposed Home Location Detection Method

The primary goal of this model is to efficiently detect home locations using sound signals, minimizing the computational time required. To achieve this, we have introduced a novel local feature extraction-based sound classification model, which comprises three key phases:

Feature Generation: This phase leverages the presented 1D-ILQP (One-Dimensional Improved Local Quadruple Pattern) and a maximum pooling-based feature extraction model. It involves the following steps:

- Creation of decomposed sub-bands of the sound signal using maximum pooling.
- Extraction of textural features from both the raw sound signal and the decomposed sub-bands using 1D-ILQP.
- Formation of a feature vector, which serves as input for the subsequent feature selection phase.
-

Feature Selection: In this phase, NCA is employed to select the most informative features. The NCA feature selection function is applied to the feature vector, leading to the identification and retention of the top 100 features.

Classification: The last phase involves classification using various machine learning methods. Eight classifiers are used to compute the classification results. These classifiers are evaluated through a 10-fold cross-validation procedure to ensure robustness and reliability of the model.

In summary, the model's workflow can be described as follows: raw sound signals are transformed into a feature vector through the application of 1D-ILQP and maximum pooling. Subsequently, feature selection using NCA is carried out, resulting in the identification of the most relevant features. Finally, the classification of home locations is performed using a range of classifiers with 10-fold cross-validation to validate the model's effectiveness. For a visual representation of this process, please refer to Figure 2 in our introduced 1D-ILQP feature extraction-based home location detection model.

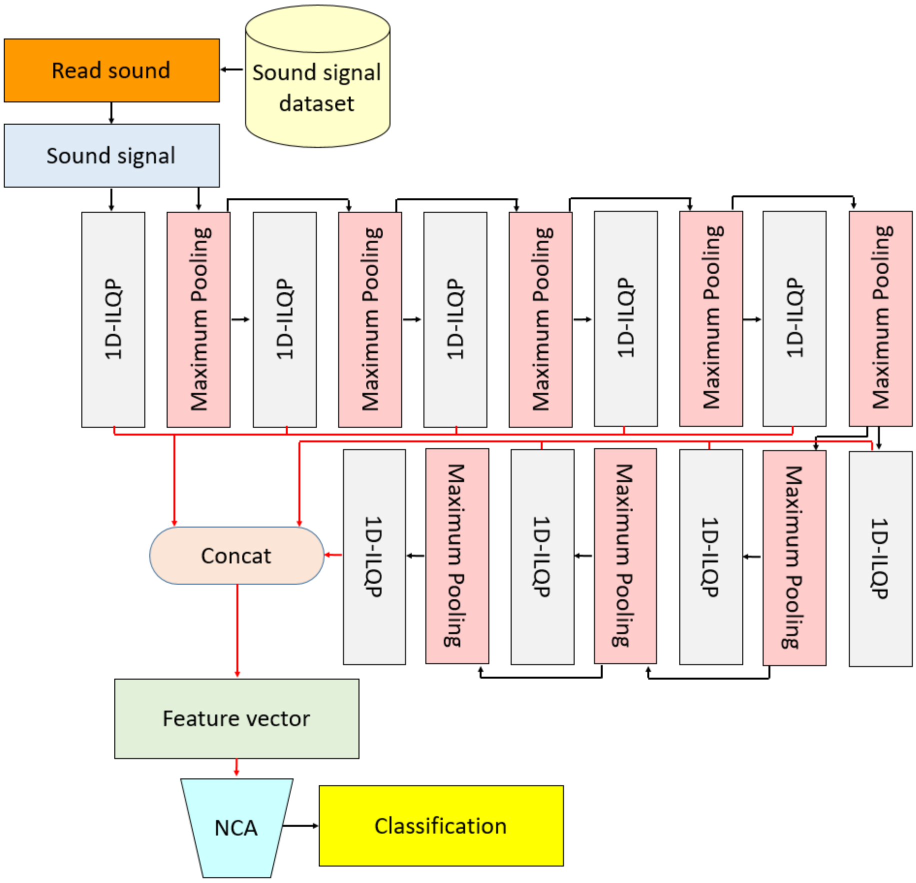

Figure 2. Block diagram of the 1D-ILQP-based sound signal classification model

This model mimics deep learning networks. It uses a manually crafted feature extractor to generate textural features. Maximum pooling is applied to create layers that function as a decomposition mechanism. These layers are instrumental in extracting high-level features. Throughout this process, the 1D-ILQP is applied nine times, extracting 768 features from each signal. During the fusion (concatenation) step, these features are combined to form a feature vector with a length of 6912 (=768×9). From this pool of 6912 features, the top 100 features are selected using the NCA selector. The results are then obtained by applying these 100 features to eight conventional classifiers.

As can be seen from Figure 1, the used feature extraction function is 1D-ILQP. The definition of the 1D-ILQP is given below. General steps of the presented 1D-ILQP based model have been listed in below.

Step 1: Create sub-bands by deploying maximum pooling. In this work, eight compressed bands have been created. The number of compressed bands have been generated by using accuracy value. The optimal accuracy score has been attained using eight compressed bands. The compressed band generation formula is denoted in below.

$$sb^1 = maxpool(sound) \quad (4)$$
$$sb^t = maxpool(sb^{t-1}), t \in \{2,3,\dots,8\} \quad (5)$$

Herein, sb defines compressed bands and $maxpool(.)$ is maximum pooling function using non-overlapping blocks with a length of two.

Step 2: Extract features from sound signal and compressed bands.

$$fv^1 = IQP(sound) \quad (6)$$
$$fv^t = IQP(sb^{t-1}), t \in \{2,3, \dots ,8\} \quad (7)$$

Herein, $IQP(.)$ defined the proposed 1D-IQP feature extraction function. In this step, the feature generation process has been defined.

Step 3: The extracted features have been merged to create final feature vector (X) with a length 6912.

$$X\big(q + 768 \times (t - 1)\big) = fv^t(q), q \in \{1,2, \dots ,768\}, t \in \{1,2, \dots ,8\} \quad (8)$$

Step 4: Choose the most discriminate/valuable 100 features from the generated 6912 features by applying NCA. NCA holds a significant place in feature selection literature and can be viewed as a feature selection variant of the k-nearest neighbors (kNN) method. NCA calculates a positive weight for each feature, with informative features receiving higher weight values and redundant features receiving lower weight values. This weighting process allows for the selection of the top 100 features based on the generated weights.

Step 5: Calculate results using the eight shallow classifiers with 10-fold cross-validation, it's important to understand the attributes of each classifier. Here's a summary of the attributes for each of the mentioned classifiers:

Decision Tree (DT): Decision trees are a non-linear model that uses a tree structure to make decisions. They split the dataset into subsets based on the most significant attribute at each node.

Linear Discriminant (LD): Linear discriminant analysis is a method used for dimensionality reduction and classification. It projects data into a lower-dimensional space while maximizing the separation between classes.

Quadratic Discriminant (QD): Quadratic discriminant analysis is a variant of linear discriminant analysis but assumes that each class has its own covariance matrix.

Naïve Bayes (NB): Naïve Bayes is a probabilistic classification algorithm based on Bayes' theorem. It assumes that features are conditionally independent.

Support Vector Machine (SVM): SVM is a powerful classification method that finds a hyperplane that best separates data points into different classes while maximizing the margin between the classes.

k-Nearest Neighbor (kNN): kNN is a simple classification algorithm that assigns a class label based on the majority class among its k-nearest neighbors in feature space.

Bagged Tree (BT): Bagging is an ensemble method that combines the predictions of multiple decision trees to reduce variance and improve accuracy.

Artificial Neural Network (ANN): Artificial neural networks are a family of machine learning models inspired by the human brain. They consist of interconnected nodes and layers and can be used for various classification tasks.

With the attributes of these classifiers in mind, you can now proceed to calculate results using each of them in your 10-fold cross-validation framework.

Table 2. Attributes of the classifiers

Classifier	Hyperparameters	Classifier	Hyperparameters
DT	Maximum number of splits: 100 Split criterion: Gini Surrogate decision split: Off	SVM	Kernel: Quadratic Box constraint level: 1 Kernel scale: Auto Multiclass method: One-vs-One
LD	Covariance structure: Full	kNN	k: 1 Distance metric: City block Weight: none
QD	Covariance structure: Full	BT	Ensemble: Bag Learner: DT Maximum number of splits: 1000 Number of learners: 30 Learning rate: 0.1
NB	Kernel: Gaussian Support: Unbounded	ANN	Number of fully connected layers: 1 Layer sizes: 100, 10, 10 Activation: ReLu Iteration limit: 1000 Lambda: 0

Results have been obtained by employing these eight shallow classifiers, with their respective hyperparameters as listed in Table 2.

5. Results

This section presents the experiments conducted in this research. The 1D-ILQP-based model was implemented in the MATLAB environment to evaluate the home location dataset used in this study. The dataset comprises 4000 sounds categorized into eight rooms, ensuring a balanced distribution. These sound samples were sourced from YouTube. The MATLAB implementation involves the utilization of several functions, including maximum pooling, 1D-ILQP, NCA, and a primary 'main' function.

To assess the model's performance, various metrics such as accuracy, precision, and F1-score were computed for each of the eight classifiers used in the study. Consequently, eight sets of results were generated. Furthermore, class-specific results for each classifier were also calculated. Initially, the comprehensive classification outcomes for each classifier are presented in Table 3.

Table 3. Overall classification performance results (%) of the used eight classifiers.

Classifier	Accuracy	Precision	F1
DT [49]	89.50	89.50	89.50
LD [50]	92.65	92.88	92.60
QD [51]	98.98	98.98	98.97
NB [52]	83.10	85.14	83.15
SVM [53]	99.68	99.68	99.68
kNN [54]	99.75	99.75	99.75
BT [55]	97.48	97.48	97.47
ANN [56]	98.95	98.95	98.95

Moreover, category-wise results have been denoted in Figure 3.

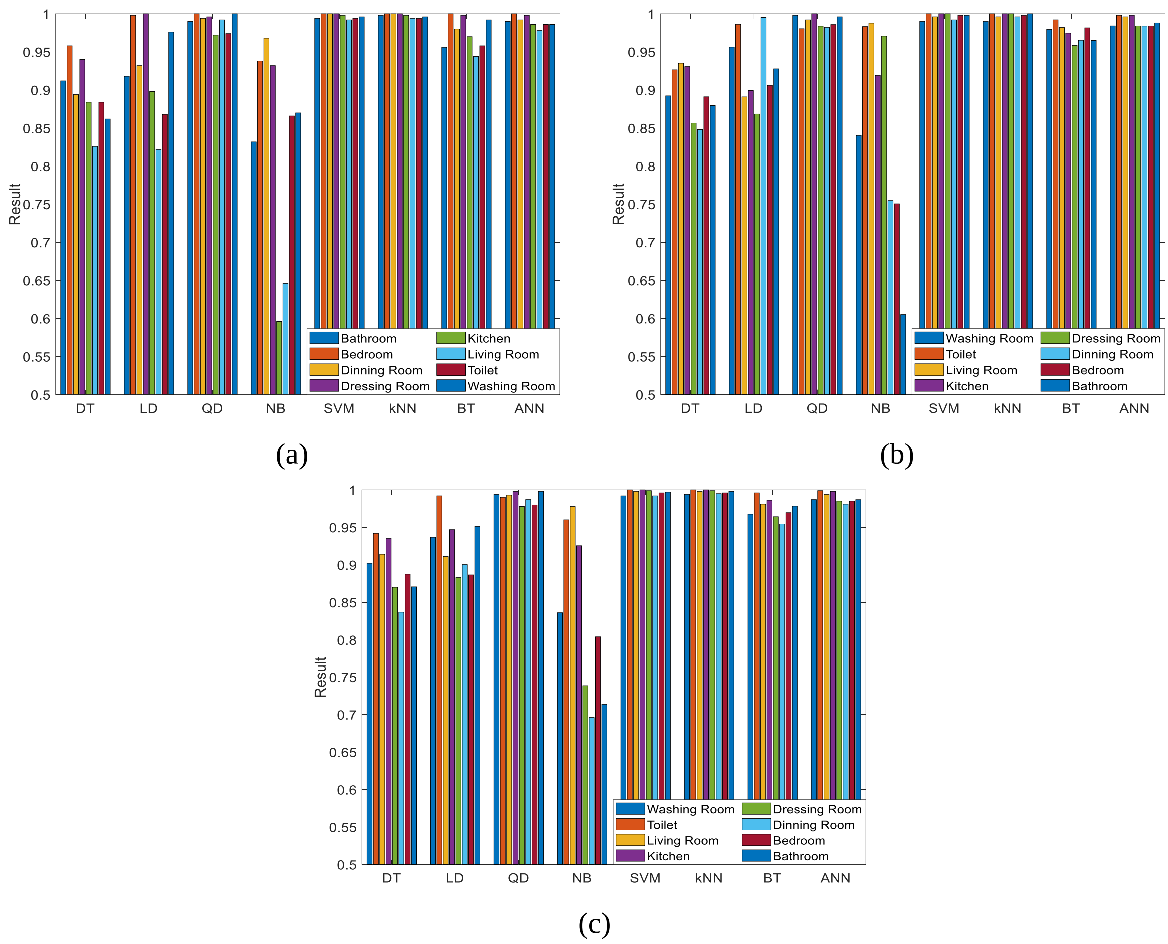

Figure 3. Category-wise classification accuracies according to the used classifiers.

Furthermore, the time complexity of the proposed 1D-ILQP pattern based sound classification has been calculated. In this model, the presented 1D-ILQP is a textural feature extractor and the time complexity of this function is equal to $O(n)$. To generate features at high levels, a simple decomposition model which is maximum pooling has been used. Maximum pooling halved the length of the used sound signal in each level. Therefore, the time burden of the presented feature extraction model is calculated as $O(nlogn)$. Moreover, we have used NCA feature selector. This feature selector is a simple feature selection function. In the classification phase, a shallow classifier has been employed to get results. In this respect, the time burden of this hand-crafted features based sound classification model is $O(nlogn + k + c)$. Herein, n is length of the sound signal, k is coefficient of NCA and c defines coefficient of the used classifier. According to this result, this model has linear time complexity.

6. Discussion

The primary objective of this research is to classify human activities within the home environment and determine the specific location within the home using sound data. This research introduces two noteworthy contributions: the compiled sound dataset and the novel 1D-ILQP function. The utilization of 1D-ILQP has facilitated the development of a novel classification model, inspired by Convolutional Neural Networks (CNNs). As previously mentioned, the presented feature extractor primarily generates low-level features. To enhance the feature extraction capabilities of 1D-ILQP, we introduce levels created through maximum pooling and subsequently generate decomposed/compressed sound signals. Our feature generator extracts feature from each of these compressed sound signals, resulting in multi-level feature extraction. An effective machine learning model must incorporate a feature extraction function to select the most valuable features and reduce the computational load on the classifier.

In this research, the Neighborhood Component Analysis (NCA) feature selector is employed, which selects the top 100 features from the 6912 initially used features. To evaluate the classification performance of the 1D-ILQP and NCA-based feature creation model, eight shallow classifiers are employed, yielding eight distinct results. The least performing classifier is Naïve Bayes (NB), which achieved an accuracy of 83.10% on the dataset. In contrast, the most successful classifier is k-Nearest Neighbor (kNN), attaining an accuracy of 99.75% on the same dataset. Additionally, the results are categorized by class for each classifier, and a comparative analysis is presented in Table 4.

Table 4. Comparative results to sound based human activity and location recognition/classification

Work and Year	Aim	Method	Number of Categories	Accuracy (%)
Mesaros et al. [47], 2010	Event detection	Hidden markov models	61 events	24
Tejada et al. [44], 2018	Indoor location estimation using activity	Contextual information extracted	10 activities, 4 locations	95
Wang et al. [46], 2019	Indoor human activity recognition	SVM model with LPCC feature	5 activities	78
Jung and Chi [43], 2020	Human activity classification	CNN	10 activities	87.6
Do et al. [45], 2021	Human activity monitoring	A two-level dynamic Bayesian network	12 activities, 6 locations	92.41
Our Work	Indoor home location classification	1D-ILQP	26 activities, 8 locations	99.75

According to Table 4, our proposed model attained 99.75% classification accuracy for eight locations classification. We proposed a hand-crafted features based model and the most of the previously presented sound classification methods used hand-crafted models except for Jung and Chi [43] method. Jung and Chi [43] used a CNN classification based model. They extracted spectrogram images from each sound and used these images as input for a residual network to compute results. In their study, they achieved an accuracy of 87.6% for a dataset with 10 classes. In contrast, our model is a manually designed learning approach, the 1D-ILQP-based model, which achieved a significantly higher classification accuracy of 99.75% for eight categories.

The benefits of the presented model are;
<u>Advantages:</u>

- The 1D-ILQP-based sound classification model achieved an impressive classification accuracy of 99.75% for eight categories, demonstrating the effectiveness of this novel approach in accurately detecting home locations using sound signals.
- The comprehensive evaluation of the model employed eight shallow classifiers. The classification results illustrate that the model's performance can vary based on the choice of classifier. The highest classification accuracy, 99.75%, was achieved using the k-Nearest Neighbor (kNN) classifier, while the Naïve Bayes (NB) classifier exhibited the lowest accuracy at 83.10%.
- The introduction of the 1D-ILQP function for feature extraction is a significant contribution. This function, when combined with maximum pooling and multi-level feature extraction, enhances the model's ability to generate informative features, improving the classification performance.

- The research utilized a balanced dataset comprising 4000 sound samples across eight categories, ensuring that each category was adequately represented. This balanced dataset contributes to the robustness and reliability of the model.
- The application of the NCA feature selector aids in selecting the most relevant features from a large set of 6912 features, optimizing the model's efficiency and effectiveness.

These findings highlight the strengths and weaknesses of the research, showcasing the promise of the 1D-ILQP-based model for indoor activity and location detection while acknowledging its specific limitations and areas for potential improvement.

7. Limitations and Future Works

In this section, the limitations of this research and future studies are presented.

The limitations of the research are presented below

- The research focuses on classifying activities and detecting locations within a home environment. The model's applicability may be limited to indoor settings and may not generalize well to other environments or broader contexts.
- While the 1D-ILQP function and feature extraction techniques are effective, manual feature engineering may not be the most scalable approach for more extensive datasets or evolving research needs.
- The model's performance is based on offline analysis, and its real-time applicability in home automation or security systems may require further development and optimization.
-

Future studies of the research are presented below

- Future work could focus on developing a real-time application of the 1D-ILQP-based sound classification model. This would involve optimizing the model's performance for immediate and continuous monitoring of activities and locations within a home environment, which could have applications in home automation and security systems.
- Extending the research to encompass a wider range of environments beyond the home, such as offices, public spaces, or outdoor settings, could expand the model's applicability. Addressing the challenges of noisy and dynamic environments presents an exciting research opportunity.
- Collecting and utilizing more extensive and diverse sound datasets would enable the model to generalize better and accommodate a broader spectrum of activities and locations. A larger dataset could also allow for more advanced machine learning techniques, such as deep learning models.
- Implementing an online learning framework could enable the model to adapt and improve its performance over time. This could be especially beneficial in environments where activities and sounds change or evolve continuously.
- Investigating how this technology can enhance human-computer interaction and user experience within smart homes or assistive technology settings is another exciting area. This research could focus on developing intuitive interfaces and control mechanisms based on sound classification.
- For applications in resource-constrained environments, such as IoT devices, research could concentrate on developing energy-efficient hardware solutions that can execute the sound classification model without significant power consumption.
- Further research into the modeling of human behavior patterns using sound signals could have implications in fields such as psychology, sociology, and healthcare. Understanding and predicting human behaviors can provide valuable insights.
-

These future works can advance the research in sound-based activity and location classification, offering innovative solutions for a wide range of applications and contributing to the evolving field of audio-based artificial intelligence. Furthermore, we are outlined our primary future objective in Figure 4.

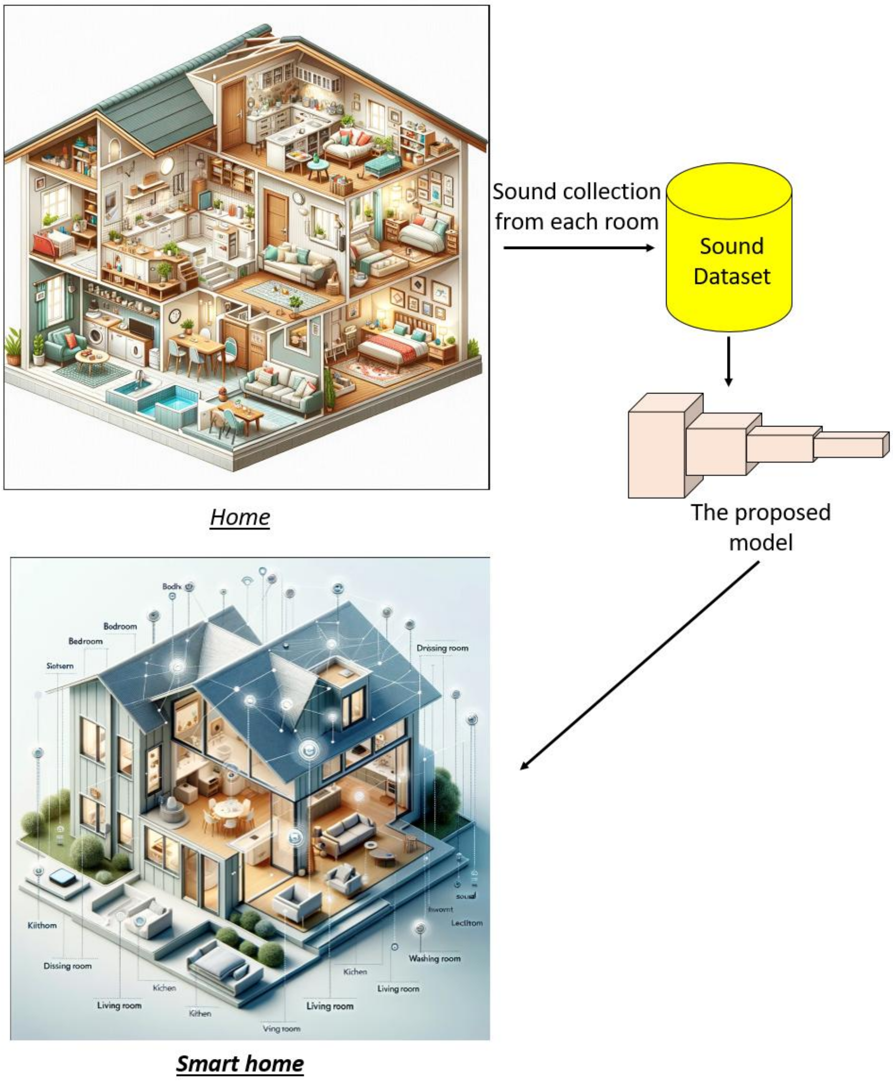

Figure 4. Converting to a smart home by implementing the proposed model

Figure 4 clearly demonstrates that the proposed model can be utilized in sound-based smart home applications. Furthermore, we can employ this model for digital forensics, particularly in sound forensics, as well as for cybersecurity and cybercrime analysis.

8. Conclusions

In this study, we presented a novel sound classification model, the 1D-ILQP-based model, aimed at classifying human activities within the home environment and accurately detecting the location of these activities using sound signals. The research has yielded significant insights and achieved remarkable results, which are summarized below:

The 1D-ILQP-based model demonstrated outstanding performance, achieving a classification accuracy of 99.75% for eight distinct categories, representing different rooms within a home. This exceptional accuracy underscores the model's ability to precisely identify activities and locations based on sound signals.

This research introduced an innovative approach to sound-based activity and location classification within a home environment. The exceptional classification accuracy, unique feature extraction techniques, and the robustness of the model are indicative of its potential for applications in home automation, security, and beyond.

As we move forward, future research will explore real-time implementation, scalability to diverse environments, and ethical considerations associated with sound-based monitoring. These endeavors aim to further enhance the applicability and responsible use of sound classification technology.

9. Author Contribution Statement

In the study, Author 1 and Author 2 contributed to the formation of the idea, design, literature review, evaluation of the results obtained, procurement of the materials used and examination of the results, Author 3 contributed to the writing of the original draft, methodology, visualization and experiments. Author 4 and Author 5, conceptualized the study and was involved in writing, reviewing and editing.

10. Ethics Committee Approval and Conflict of Interest Statement

There is no need to obtain ethics committee permission for the prepared article. There is no conflict of interest with any person/institution in the prepared article.

11. References

[1] C. A. Ronao and S.B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," Expert Syst. Appl., vol. 59, pp. 235-244, 2016.

[2] B. Dong and K. P. Lam, "Building energy and comfort management through occupant behaviour pattern detection based on a large-scale environmental sensor network," J. Build. Perform. Simul., vol. 4, pp. 359-369, 2011.

[3] J. G. Ortega, L. Han, N. Whittacker, and N. Bowring, "A machine-learning based approach to model user occupancy and activity patterns for energy saving in buildings," Sci. Inf. Conf. (SAI), IEEE, pp. 474-482, 2015.

[4] A. Khosrowpour, J. C. Niebles, and M. Golparvar-Fard, "Vision-based workface assessment using depth images for activity analysis of interior construction operations," Autom. Constr., vol. 48, pp. 74-87, 2014.

[5] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang, "Convolutional neural networks for human activity recognition using mobile sensors," in 6th Int. Conf. Mob. Comput. Appl. Serv., IEEE, pp. 197-205, 2014.

[6] A. Jalal, Y.-H. Kim, Y.-J. Kim, S. Kamal, and D. Kim, "Robust human activity recognition from depth video using spatiotemporal multi-fused features," Pattern Recognit., vol. 61, pp. 295-308, 2017.

[7] S. Kamal, A. Jalal, and D. Kim, "Depth images-based human detection, tracking and activity recognition using spatiotemporal features and modified HMM," J. Electr. Eng. Technol., vol. 11, pp. 1857-1862, 2016.

[8] A. Franco, A. Magnani, and D. Maio, "A multimodal approach for human activity recognition based on skeleton and RGB data," Pattern Recognit. Lett., 2020.

[9] M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," Future Gener. Comput. Syst., vol. 81, pp. 307-313, 2018.

[10] Y. Chen and C. Shen, "Performance analysis of smartphone-sensor behavior for human activity recognition," IEEE Access, vol. 5, pp. 3095-3110, 2017.

[11] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," Expert Syst. Appl., vol. 105, pp. 233-261, 2018.

[12] E. Lattanzi, L. Calisti, and P. Capellacci, "Lightweight accurate trigger to reduce power consumption in sensor-based continuous human activity recognition," Pervasive Mob. Comput., 2023.

[13] H. Feng, Q. Shen, R. Song, L. Shi, and H. Xu, "ATFA: Adversarial Time–Frequency Attention network for sensor-based multimodal human activity recognition," Expert Syst. Appl., vol. 236, 2024.

[14] Z. Chen, Q. Zhu, Y. C. Soh, and L. Zhang, "Robust human activity recognition using smartphone sensors via CT-PCA and online SVM," IEEE Trans. Ind. Inf., vol. 13, pp. 3070-3080, 2017.

[15] A. Ignatov, "Real-time human activity recognition from accelerometer data using Convolutional Neural Networks," Appl. Soft Comput., vol. 62, pp. 915-922, 2018.

[16] Z. Qin, Y. Zhang, S. Meng, Z. Qin, and K.-K. R. Choo, "Imaging and fusing time series for wearable sensor-based human activity recognition," Inf. Fusion, vol. 53, pp. 80-87, 2020.

[17] F. S. Abuhoureyah, Y. C. Wong, and A. S. B. M. Isira, "WiFi-based human activity recognition through wall using deep learning," Eng. Appl. Artif. Intell., vol. 127, 2024.

[18] A. Mastakouris, G. Andriosopoulou, D. Masouros, P. Benardos, G.-C. Vosniakos, and D. Soudris, "Human worker activity recognition in a production floor environment through deep learning," J. Manuf. Syst., vol. 71, pp. 115-130, 2023.

[19] M. N.U. Hasan, and C. R. Stannard, "Exploring online consumer reviews of wearable technology: The Owlet Smart Sock," Res. J. Text. Apparel, 2022.

[20] M. P. Buttner and L. D. Stetzenbach, "Monitoring airborne fungal spores in an experimental indoor environment to evaluate sampling methods and the effects of human activity on air sampling," Appl. Environ. Microbiol., vol. 59, pp. 219-226, 1993.

[21] A. Bansal and N. K. Garg, "Environmental Sound Classification: A descriptive review of the literature," Intell. Syst. Appl., 2022.
[22] Y. Chen, Q. Guo, X. Liang, J. Wang, and Y. Qian, "Environmental sound classification with dilated convolutions," Appl. Acoust., vol. 148, pp. 123-132, 2019.
[23] S. Abdoli, P. Cardinal, and A. L. Koerich, "End-to-end environmental sound classification using a 1D convolutional neural network," Expert Syst. Appl., vol. 136, pp. 252-263, 2019.
[24] A. Khamparia, D. Gupta, N. G. Nguyen, A. Khanna, B. Pandey, and P. Tiwari, "Sound classification using convolutional neural network and tensor deep stacking network," IEEE Access, vol. 7, pp. 7717-7727, 2019.
[25] A. Bansal and N. K. Garg, "Environmental Sound Classification using Hybrid Ensemble Model," Procedia Comput. Sci., vol. 218, pp. 418-428, 2023.
[26] S. Dong, Z. Xia, X. Pan, and T. Yu, "Environmental sound classification based on improved compact bilinear attention network," Digit. Signal Process., vol. 141, 2023.
[27] M. YILDIRIM, " Automatic classification of environmental sounds with the mfcc method and the proposed deep model," Fırat Univ. J. Eng. Sci., vol. 34, pp. 449-457, 2022.
[28] A. Baró, P. Riba, J. Calvo-Zaragoza, and A. Fornés, "From optical music recognition to handwritten music recognition: A baseline," Pattern Recognit. Lett., vol. 123, pp. 1-8, 2019.
[29] H. Tang, and N. Chen, "Combining CNN and Broad Learning for Music Classification," IEICE Trans. Inf. Syst., vol. 103, pp. 695-701, 2020.
[30] D. Snyder, D. Garcia-Romero, G. Sell, A. McCree, D. Povey, and S. Khudanpur, "Speaker recognition for multi-speaker conversations using x-vectors," ICASSP 2019 - IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), pp. 5796-5800, 2019.
[31] J. Villalba, N. Chen, D. Snyder, D. Garcia-Romero, A. McCree, G. Sell, J. Borgstrom, F. Richardson, S. Shon, F. Grondin, "State-of-the-art speaker recognition for telephone and video speech: The JHU-MIT submission for NIST SRE18," Proc. Interspeech, pp. 1488-1492, 2019.
[32] J. Nicholson, K. Takahashi, and R. Nakatsu, "Emotion recognition in speech using neural networks," Neural Comput. Appl., vol. 9, pp. 290-296, 2000.
[33] Y.S. Seo, and J.-H. Huh, "Automatic emotion-based music classification for supporting intelligent IoT applications," Electron., vol. 8, p. 164, 2019.
[34] M. Yildirim, "Automatic diagnosis of snoring sounds with the developed artificial intelligence-based hybrid model," Turk. J. Sci. Technol., vol. 17, pp. 405-416, 2022.
[35] N. Takahashi, M. Gygli, B. Pfister, and L. Van Gool, "Deep convolutional neural networks and data augmentation for acoustic event detection," arXiv preprint arXiv:1604.07160, 2016.
[36] X. Xia, R. Togneri, F. Sohel, and D. Huang, "Random forest classification based acoustic event detection utilizing contextual-information and bottleneck features," Pattern Recognit., vol. 81, pp. 1-13, 2018.
[37] X. Xia, R. Togneri, F. Sohel, and D. Huang, "Auxiliary classifier generative adversarial network with soft labels in imbalanced acoustic event detection," IEEE Trans. Multimedia, vol. 21, pp. 1359-1371, 2018.
[38] F. Saki et al., "Open-set evolving acoustic scene classification system," 2019.
[39] T. Doan, H. Nguyen, D. T. Ngo, L. Pham, and H. H. Kha, "Acoustic scene classification using a deeper training method for convolution neural network," 2019 Int. Symp. Electr. Electron. Eng. (ISEE), pp. 63-67, 2019.
[40] J. Xie, and M. Zhu, "Investigation of acoustic and visual features for acoustic scene classification," Expert Syst. Appl., vol. 126, pp. 20-29, 2019.
[41] Y. Han, J. Park, and K. Lee, "Convolutional neural networks with binaural representations and background subtraction for acoustic scene classification," Detec. and Classif. of Acou. Sce. and Events (DCASE), pp. 1-5, 2017.
[42] L. Yang, L. Tao, X. Chen, and X. Gu, "Multi-scale semantic feature fusion and data augmentation for acoustic scene classification," Appl. Acoust., vol. 163, p. 107238, 2020.
[43] M. Jung, and S. Chi, "Human activity classification based on sound recognition and residual convolutional neural network," Autom. Constr., vol. 114, p. 103177, 2020.
[44] C.E. Galván-Tejada, F. López-Monteagudo, O. Alonso-González, J.I. Galván-Tejada, J.M. Celaya-Padilla, H. Gamboa-Rosales, R. Magallanes-Quintanar, L.A. Zanella-Calzada, "A generalized

model for indoor location estimation using environmental sound from human activity recognition," ISPRS Int. J. Geo-Inf., vol. 7, p. 81, 2018.
[45] H. M. Do, K. C. Welch, and W. Sheng, "Soham: A sound-based human activity monitoring framework for home service robots," IEEE Trans. Autom. Sci. Eng., 2021.
[46] W. Wang, F. Seraj, N. Meratnia, and P. J. Havinga, "Privacy-aware environmental sound classification for indoor human activity recognition," Proc. 12th ACM Int. Conf. Pervasive Technol. Assist. Environ., pp. 36-44, 2019.
[47] A. Mesaros, T. Heittola, A. Eronen, and T. Virtanen, "Acoustic event detection in real life recordings," 18th Eur. Signal Process. Conf., pp. 1267-1271, 2010.
[48] J. Goldberger, G. E. Hinton, S. Roweis, and R. R. Salakhutdinov, "Neighbourhood components analysis," Adv. Neural Inf. Process. Syst., vol. 17, pp. 513-520, 2004.
[49] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," IEEE Trans. Syst. Man Cybern., vol. 21, pp. 660-674, 1991.
[50] W. Zhao, R. Chellappa, and N. Nandhakumar, "Empirical performance analysis of linear discriminant classifiers," Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit., pp. 164-169, 1998.
[51] C. E. Thomaz, D. F. Gillies, and R. Q. Feitosa, "A new quadratic classifier applied to biometric recognition," Int. Workshop Biometric Auth., Springer, pp. 186-196, 2002.
[52] A. Y. Ng and M. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes," Adv. Neural Inf. Process. Syst., pp. 841-848, 2002.
[53] V. Vapnik, "The support vector method of function estimation," Nonlinear Modeling, Springer, pp. 55-85, 1998.
[54] L. E. Peterson, "K-nearest neighbor," Scholarpedia, vol. 4, p. 1883, 2009.
[55] T. Hothorn and B. Lausen, "Bagging tree classifiers for laser scanning images: A data-and simulation-based strategy," Artif. Intell. Med., vol. 27, pp. 65-79, 2003.
[56] Q. K. Al-Shayea, "Artificial neural networks in medical diagnosis," Int. J. Comput. Sci. Issues, vol. 8, pp. 150-154, 2011.