World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Research Article)

Check for updates

Machine learning-based equipment sound classification for advanced construction management and site supervision

Suat Gokhan Ozkaya [1], Mehmet Baygin [2,*], Sengul Dogan [3] and Turker Tuncer [3]

[1] *Civil Technology, Ardahan Vocational School of Technical Sciences, Ardahan University, Ardahan, Turkey.*
[2] *Department of Computer Engineering, Faculty of Engineering and Architecture, Erzurum Technical University, Erzurum, Turkey.*
[3] *Department of Digital Forensics Engineering, College of Technology, Firat University, Elazig, Turkey.*

World Journal of Advanced Research and Reviews, 2025, 26(03), 317–329

Publication history: Received on 28 April 2025; revised on 31 May 2025; accepted on 03 June 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.26.3.2178

Abstract

This study focuses on machine learning classification of the sounds of equipment operating in the construction site environment to support improved construction management and site supervision processes. The research utilizes a large, openly available, open access audio dataset of seven different types of equipment, collected in the field under real conditions in urban regeneration projects initiated after a major earthquake in Elazığ. The dataset consists of 15,588 sounds recorded from vehicles such as bulldozers, excavators, dump trucks, graders, loaders, mixer trucks and rollers used on the construction site.

In the developed classification system, discriminative features were first extracted from equipment sounds using Local Binary Patterns (LBP) and statistical moments. In the feature selection stage, the Neighborhood Component Analysis (NCA) and Chi-Square (Chi2) method is applied to identify the most significant features and dimensionality reduction is achieved. In the final stage, Support Vector Machines (SVM) and k-Nearest Neighbor (kNN) classifiers are used to discriminate the equipment types with high accuracy. The findings show that the proposed method makes a significant contribution to construction management objectives such as effective monitoring of vehicles and equipment on the construction site, resource management and process tracking. In addition, the transparency and reproducibility provided by the open dataset provides a strong basis for further studies in the related field.

Keywords: Local Binary Pattern; Signal Processing; Statistical Moment; Machine Learning; Construction Management

1. Introduction

Nowadays, construction projects require more effective management approaches and innovative technologies as they increase in scope and complexity [1]. Construction management undertakes multi-dimensional responsibilities such as resource management, process monitoring, occupational health and safety to ensure that projects are completed on time, on budget and to high quality standards [2]. Accordingly, effective and accurate monitoring of activities taking place at the construction site is critical for both reducing costs and increasing occupational safety [3].

In recent years, artificial intelligence and machine learning-based approaches have attracted increasing attention in civil engineering and construction management [4]. Especially by analyzing large data sources (such as images, sensors, sound) on construction sites, it is possible to provide faster, objective and sustainable solutions compared to traditional methods. Automatic classification of audio signals is one of the innovative application areas emerging in this context [5]. By analyzing the characteristic sounds produced by vehicles and equipment on the construction site, it is possible

* Corresponding author: Mehmet Baygin.

to obtain information about which type of equipment is operating when and for how long; thus, managerial functions such as progress tracking, resource utilization analysis and detection of potential safety risks can be digitized and automated [6].

Construction site monitoring processes, which are traditionally carried out through manual observation and reporting, create various difficulties in the construction industry due to their time-consuming and human error-prone nature [3, 7]. However, thanks to AI-powered sound classification systems, it is possible to automatically monitor and report on equipment in the field, significantly increasing the transparency and traceability of processes [8]. In addition, sound data can be collected with low-cost microphones and portable devices, enabling effective inspection over large areas even in large-scale projects [9].

As a result, the use of artificial intelligence-based audio classification technologies in construction management and civil engineering offers significant advantages in terms of process efficiency, occupational health and safety, resource optimization and sustainability [10]. In order to demonstrate the application potential of these technologies, this study performs equipment classification with machine learning on audio data collected in a real construction site environment.

2. Literature review

Nowadays, artificial intelligence-based methods are actively used for classification in many different disciplines [11]. These approaches, especially for text, signal and image processing, are very successful [12]. In this study, automatic sound classification of construction equipment working on the construction site has been provided for construction management, which is a very important issue in civil engineering [13]. This approach, which can be used especially for autonomous monitoring of construction sites, has a lightweight architecture. In this context, similar studies in the literature are summarized in Table 1.

Table 1 Construction site audio classification in the literature

Study and Year	Dataset	Method	Results	Limitation(s)
Scarpiniti et al., 2021 [13]	~10 types; private dataset	DBN + spectral audio features	~98% accuracy	First deep model for this domain; very high accuracy; not public
Maccagno et al., 2021 [14]	5 types; private dataset	8-layer CNN on Mel-spectrogram	~97% accuracy	CNNs succeed on construction audio; limited classes/dataset
Sherafat et al., 2021 [15]	Synthetic & limited real; private	CNN, 2-level, multi-label (mixes)	Detected multiple overlapping machines	Synthetic audio allows training; less robust on real noise
Akbal & Tuncer, 2022 [16]	Large custom, private	256 features + SVM	~99.4% accuracy	Classic ML, high accuracy; likely overfit, not public
Akbal et al., 2022 [17]	15,588 samples, 7 types, private	DES S-box, tent pooling, kNN/SVM	~97% accuracy	Lightweight, big data, high perf.; custom feature gen, not public
Sherafat et al., 2022 [18]	Synthetic+real, 16 classes	CNN multi-label, mixes	F1~90% multi-label	Detects multiple machines; complex, less robust on real noise
Xiong et al., 2022 [19]	8 classes, real (private)	5-layer CNN+Bi-GRU (CRNN)	F1~91.5%; high precision	Multi-label, pre-trained, robust; weak on intermittent, not public
Jeong et al., 2022 [20]	820 videos, auto, noisy	ML+CV auto-labeling, 6 classifiers	64-93% accuracy	Novel auto-labeled data; noisy, label errors
Jeong, Park & Ahn, 2025 [21]	Large, vision-labeled, private	CV+audio, SVM	F1: 61–91% (type), 52–87% (action)	Multi-modal, robust for types; harder for actions
Scarpiniti et al., 2023 [22]	5 classes, private	Leaky Echo State Net (ESN)	~95.3% accuracy	ESN near CNN performance; simple, but less general

3. Motivation and Our Method

The increasing complexity and size of today's construction projects have made the need for more efficient, traceable and automated solutions in construction management processes greater than ever [3]. Monitoring vehicle and equipment usage on construction sites is a critical element for improving resource management, increasing occupational safety and digitizing the process [23]. However, traditional methods of manual monitoring and reporting do not provide an effective solution as they are time-consuming and prone to human error [7]. Therefore, approaches based on automatic sound classification are emerging as an important innovation in construction management and site surveillance.

In this study, a novel multi-stage model is developed for automatic classification of equipment sounds recorded in a construction site environment. In the proposed approach, Local Binary Patterns (LBP) [24] and statistical moments [25] are used for feature extraction from equipment sounds. These two methods have the capacity to effectively represent the characteristic sound features of the equipment in both temporal and frequency terms. The resulting feature vectors are then given as input to the Neighborhood Component Analysis (NCA) [26] and Chi-Square (Chi2) [27] algorithms separately in the feature selection phase. This resulted in two separate feature vectors produced by two different feature selection algorithms, rich in information and reduced in size.

Each selected feature vector is trained and evaluated separately with both k-Nearest Neighbor (kNN) [28] and Support Vector Machines (SVM) [29] classifiers. In this way, the performance of different combinations of feature selection and classification is thoroughly analyzed. In the last stage, the majority voting method was applied to generalize the four different classification results obtained. This method improves the final classification performance by taking into account the unanimous results of all models.

This multi-stage and modular approach aim to increase the generalizability and robustness of the model while enabling highly accurate identification of various equipment sounds recorded under different conditions at the construction site. Thus, it is aimed to provide an effective solution to many critical needs in construction management processes such as automatic and real-time equipment monitoring, optimization of resource utilization and increasing occupational safety.

3.1. Innovations and Contributions

The main original contributions and innovative aspects of this study to the civil engineering and construction management literature are summarized below:

- The study was conducted on a comprehensive, open-access sound dataset belonging to seven different types of equipment in a completely real construction site environment. This ensures that the model is resilient to the challenges and realistic variations that may be encountered in field conditions.
- In the proposed method, the feature vectors obtained in the feature extraction stage were processed separately using two different feature selection algorithms (NCA and Chi2). Thus, the meaningful features selected by different algorithms were analyzed comparatively, and the generalizability of the model was increased.
- Testing different classifier algorithms such as kNN and SVM separately on each of the two selected feature vectors and performing the final classification using majority voting is an important approach that increases the reliability and overall performance of the model.
- Compared to deep learning-based heavy models, a lightweight and computationally efficient method based on feature engineering has been developed, providing a low-cost and field-practical solution capable of handling big data.
- The developed model enables automatic and highly accurate tracking of equipment on construction sites, thereby addressing critical management needs such as optimization of resource utilization, process tracking, work safety, and increased project efficiency in construction management processes.
- The modularity of the methods and analyses used in the study provides a structure that can be easily adapted to different data sets, additional equipment types, or new feature extraction/selection algorithms. Furthermore, the use of an open-access data set fills an important gap in the literature in terms of reproducibility and transparency.

3.2. Dataset

In this study, the developed model is tested on an openly available audio dataset [17]. These sounds, which are mainly collected using construction equipment used in construction sites, are used for automatic classification with the developed model. The dataset used in the research was collected from construction sites in Elazığ region, Turkey, and includes sounds from 7 different pieces of equipment. The data collected in the research was recorded using a mobile cell phone and has a frequency of 44.1 kHz. The collected data is about 362 minutes in total and is divided into small segments to be used in the training and testing phases of the model. In this way, a total of 15,588 short audio recordings were obtained. The dataset used in the research can be accessed from the link below. In addition, the distribution of the samples in the dataset is given in Table 2.

Access Link: https://www.kaggle.com/datasets/turkertuncer/construction-site-monitoring-sound-signals-dataset

Table 2 Equipment Classes and Sample Distribution in the Dataset

Equipment Type	Sample Size
Bulldozer	3400
Dump Truck	1270
Excavator	2994
Grader	1630
Loader	2929
Mixer Truck	1455
Cylinder	1910
Total	15588

3.3. LBPStat-SelectFusion

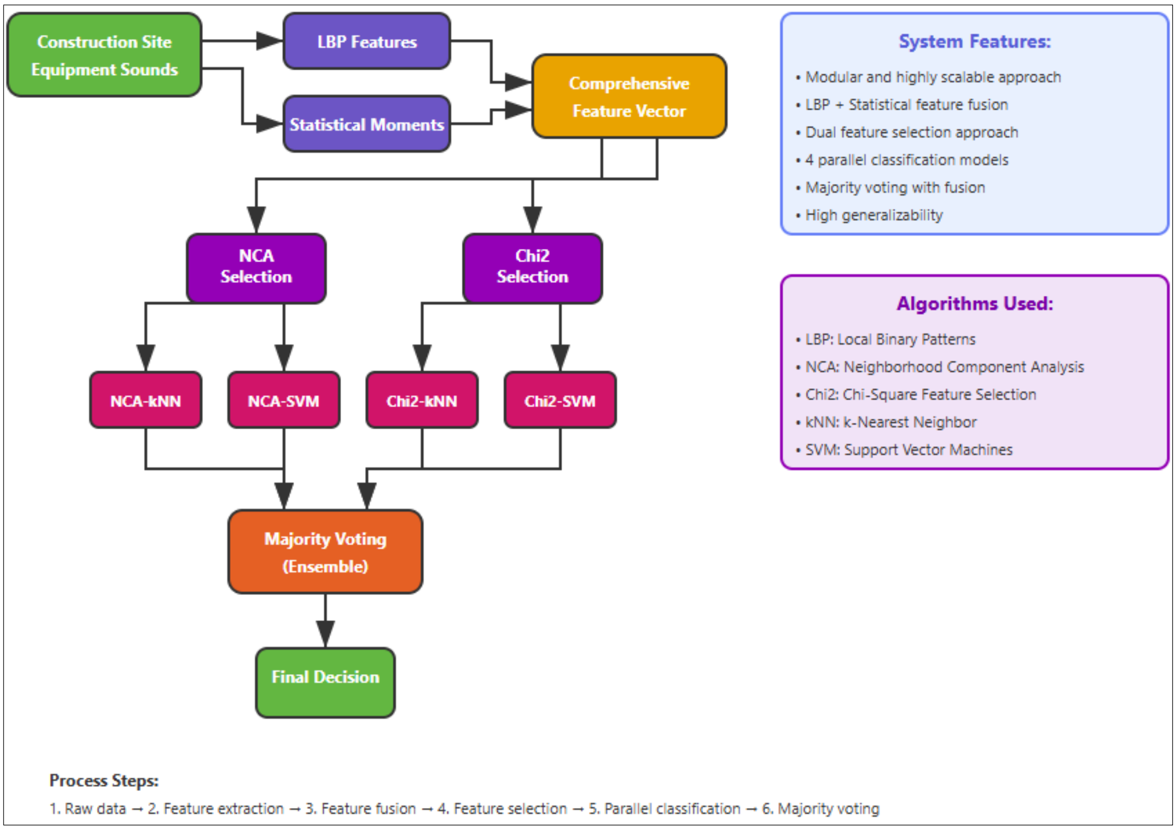


Figure 1 LBPStat-SelectFusion method

The LBPStat-SelectFusion method proposed in this study is a multi-stage, modular machine learning approach developed for the automatic classification of equipment sounds recorded at construction sites. The method is based on extracting distinctive features from sound signals through both local binary patterns (LBP) [24] and statistical moments

[25]. The resulting comprehensive feature vectors are then processed using a two-way feature selection strategy: Neighborhood Component Analysis (NCA) [26] and Chi-Square (Chi2) [27] algorithms, each producing uniquely selected sub-vectors. Each selected feature vector is then evaluated separately using both k-Nearest Neighbor (kNN) [28] and Support Vector Machines (SVM) [29] classifiers; the final decision is obtained by combining the outputs of these four parallel models through majority voting. This maximizes both the model's generalizability and classification accuracy. A block diagram summarizing the flow of the proposed LBPStat-SelectFusion method is presented in Figure 1.

Figure 1 shows the flow of the LBPStat-SelectFusion method. As can be seen in Figure 1, the system takes sound data collected from the construction site as input and extracts features from these sound files using LBP and statistical moment methods. Subsequently, these feature vectors are combined, and the resulting feature vector is sent to the feature selection phase. In this phase of the algorithm, two separate selected feature vectors are provided using the NCA and Chi2 algorithms. The selected feature vectors are then fed into the kNN and SVM algorithms, which are well-known algorithms in the literature. In this phase, a total of 4 classification results are obtained. In the final phase of the developed model, the majority voting algorithm is applied to both generalize the classification results and improve classification accuracy. As a result, the obtained classification results are combined, and the final classification results are obtained. The methods used in this developed model are detailed in the following sections.

3.4. Local Binary Pattern (LBP)

LBP method is a feature extraction approach frequently used in image processing. In this approach, which is basically based on the principle of comparing the center pixel in an image with the surrounding pixels (signum function), all pixels are compared through a sliding window and a feature vector is obtained. In this study, feature extraction is performed on 1D signals using the LBP method. Basically, in this approach where the signal is divided into overlapping blocks, firstly, 9 sequential values are taken from the signal and transformed into a 3x3 matrix. Then, the center value of this matrix is compared with the neighboring values around it and a binary array is obtained. The binary array obtained with this approach, which is mathematically expressed as a signum function, is then converted into a decimal number and placed on the histogram curve according to the value obtained. In the next step, the 2nd overlapping block is taken and the process steps are repeated. Finally, the resulting histogram curve is used as the feature vector. The steps of this approach, in which 256 features are obtained, are given below.

- A sliding window consisting of 9 samples is defined on the signal (including the sample at the center).
- For each window, the sample at the center is compared with the eight neighboring samples.
- If the neighboring value is greater than the center value, the corresponding bit is set to 1, otherwise it is set to 0, resulting in an 8-bit binary number.
- The decimal equivalent of this number (0–255) is calculated.
- The histogram of the LBP codes obtained from all windows along the signal is calculated, and this histogram is used as the LBP feature vector for the sample.

The LBP method is generally an approach used to extract features from image data. However, in this study, it was used to extract features from signal data. A block diagram of this algorithm, which essentially produces a feature vector of length 256, is shown in Figure 2.

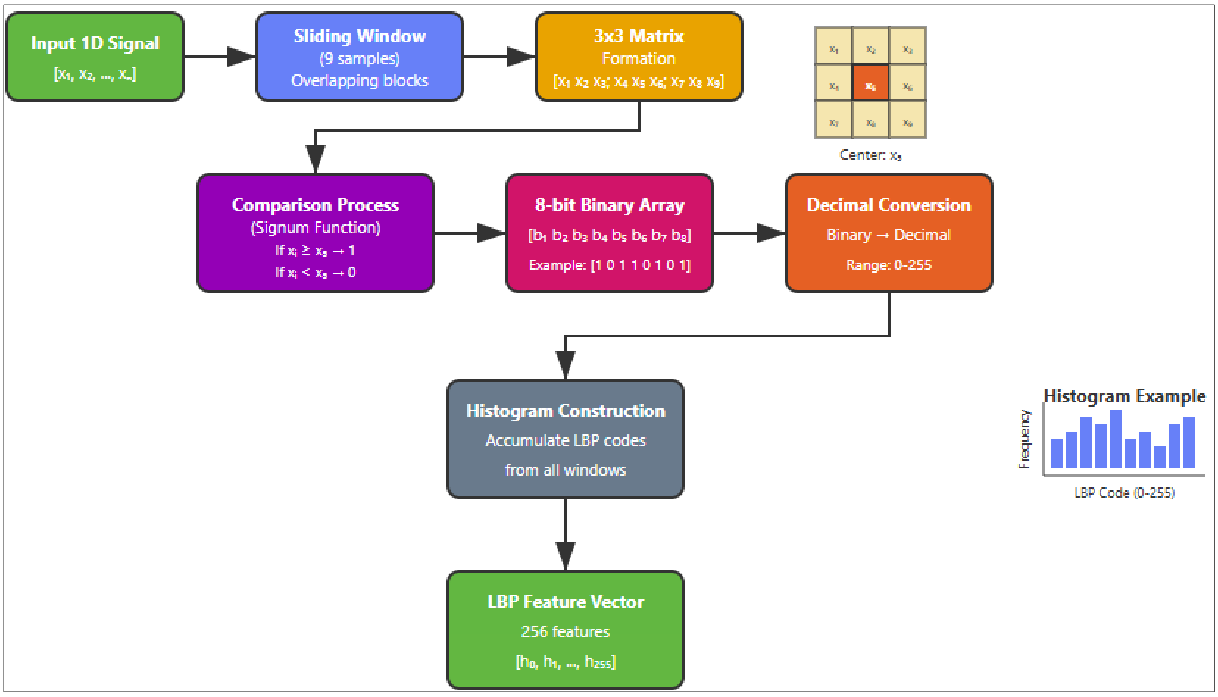


Figure 2 LBP method for 1D signal feature extraction

As shown in Figure 2, 9-sample blocks are created on the input signal; for each block, the difference between the center value (5th sample) and eight neighbors is taken, and an 8-bit LBP code is obtained as a result of thresholding. The histogram of these codes formed throughout the signal constitutes the LBP-based feature vector for that sample.

The LBP method offers a highly suitable feature extraction approach for modeling the discrete characteristics of construction equipment sounds, thanks to its low computational cost and ability to capture local structures with high resolution. This allows each sound sample to be represented by a compact, descriptive, and noise-resistant feature vector.

3.5. Statistical Moments

Statistical moments and various statistical measures are frequently used to characterize the structure of audio signals. Statistical moments are numerical values that summarize different aspects of a signal's amplitude distribution and represent the signal's fundamental statistical properties. This approach enables the distinction between sound signals from different equipment types based on the fundamental features obtained in the time domain. The statistical features used in this study are derived from both the original signal and the absolute-valued signal. The feature extraction process applied within the developed model is summarized below:

- Basic statistical measures: maximum, minimum, mean, median, variance, standard deviation, peak (max-min), RMS, total energy, mode, mean absolute deviation (MAD), and various ratios.
- High-order moments: skewness, kurtosis.
- Entropy-based measures: Features that measure the irregularity and information content of the signal (separately for both the original and absolute value signals), such as Shannon entropy, duration entropy, log energy entropy, thresholding, and norm entropies.
- Combinations and ratios: additional combinations summarizing the characteristic structures of the signal, such as median-standard deviation difference, median/standard deviation ratio, maximum amplitude, and absolute amplitudes.

A total of 40 different features have been calculated to represent both the basic statistical building blocks of the signal and its distribution and irregularity characteristics. These features provide information about both the amplitude and energy structure specific to the operation of the equipment and the complexity and variation of the signal. Statistical moments offer an effective and widely used feature extraction approach for classifying construction equipment sounds using machine learning, thanks to their low computational costs and ability to represent the general characteristics of the signal. This approach ensures the model's generalizability and successful operation under various field conditions.

3.6. Feature Selection

In machine learning-based classification problems, high-dimensional feature space can negatively affect the generalizability of the model and lead to unnecessary computational costs. Therefore, feature selection methods improve model performance and prevent unnecessary information redundancy by selecting the most meaningful and discriminative features in the dataset. In this study, two different feature selection algorithms were used. These algorithms are NCA and Chi2 algorithm, respectively. The details of these algorithms are provided in the subsections.

3.6.1. Neighborhood Component Analysis (NCA)

NCA is a supervised feature selection method that focuses on identifying feature subsets that maximize classification performance. NCA learns the projection that best separates the data in the feature space using a kNN classifier by optimizing a linear transformation matrix. Thus, the most informative features that enhance the model's classification accuracy are selected. NCA can be flexibly adapted to the dataset due to its non-parametric structure and delivers effective results, especially in complex datasets with multiple feature groups.

3.6.2. Chi-Square (Chi2)

Chi2 test is a statistical method used to evaluate the relationship between a categorical target variable and continuous or categorical attributes. In the feature selection process, the independence hypothesis between each attribute and the target variable is tested, and the discriminative power of the attribute on the target variable is measured using the calculated chi-square statistic. Features with high chi-square scores are considered more informative for classification and are selected. This method provides fast and efficient pre-selection, especially in high-dimensional feature sets.

3.7. Classification

Following the feature extraction and selection stages, two different supervised machine learning algorithms were used to classify the obtained feature vectors: kNN and SVM. Both algorithms are widely used with success in voice and time series-based classification applications in the literature. Therefore, these methods were chosen for the developed model, and 10-fold cross-validation (CV) was applied as the validation strategy.

3.7.1. k-Nearest Neighbors (kNN)

The kNN algorithm is a lightweight method that is frequently used in the literature and is especially preferred for classification problems. In this study, the kNN algorithm was applied to feature vectors selected using both NCA and Chi2. The parameters of the kNN algorithm used in the research are given below:

- Number of neighbors: 5
- Distance metric: Euclidean
- Cross validation (Kfold): 10-fold
- Distance weight: Equal
- Standardization: Not applied

3.7.2. Support Vector Machine (SVM)

Another method used in this research is the SVM algorithm. As with the kNN algorithm, this method, which is frequently preferred in the literature, produces very successful results especially in classification problems. The parameters of this method, which basically classifies the features selected with Chi2 and NCA algorithms, are given below:

- Kernel function: Linear
- Cross validation (Kfold): 10-fold
- Box constraint: 1
- Kernel scale: Auto
- Polynomial order: 3
- Majority Voting

In multi-classifier systems, ensemble methods are often preferred to improve overall model performance by combining the outputs of different feature selections and classifier combinations. In this study, four independent prediction outputs were obtained by separately evaluating the feature vectors selected by both NCA and Chi2 algorithms with kNN and SVM classifiers. The final decision was made by combining the classification results of these four parallel models. In the majority voting method, for each test example, the class labels produced by the four different models are considered, and the class with the most votes is assigned as the final decision. If any class receives an equal number of votes, a predefined preference order or random selection can be applied. This ensemble approach increases the overall system accuracy and stability by taking advantage of the accuracy of other models when a single model makes a mistake. Majority voting is an effective ensemble method that supports model generalizability, especially when there is diversity in feature selection and classifier types, and enhances system reliability in noisy or uncertain conditions. In this study, the automatic and accurate classification of construction site equipment sounds was targeted through the collective evaluation of four different models.

3.8. Experimental results

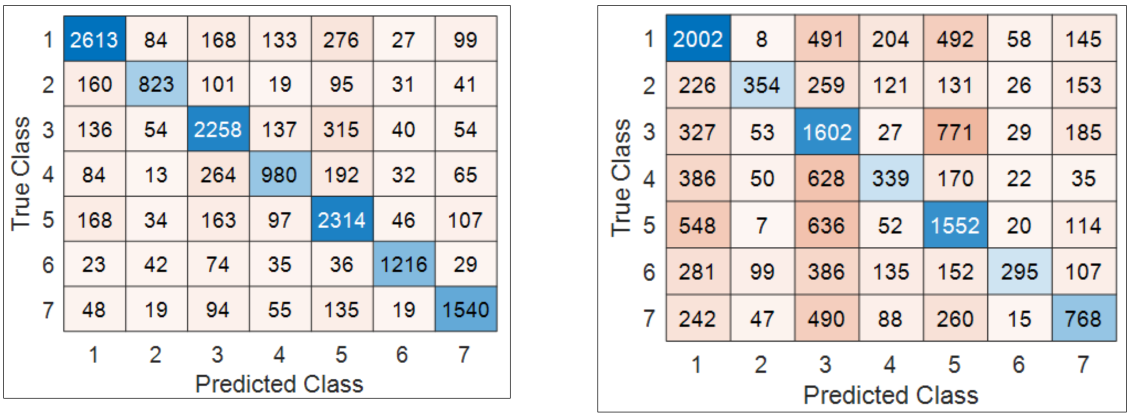
The proposed LBPStat-SelectFusion method has been comprehensively evaluated on the open-access sound dataset detailed in previous sections for the automatic classification of construction site equipment sounds. All feature extraction, selection, and classification steps of the model were implemented using the MATLAB 2021b software platform. During the experimental studies, both LBP and statistical moment-based features were extracted from the raw audio data; feature selection was then performed using NCA and Chi2 algorithms, and finally, performance comparisons were conducted using kNN and SVM classifiers. The final classification performance was obtained by combining the results of four different models using the majority voting method.

All analyses and calculations were performed on a CPU-based server. The hardware specifications of the experimental environment are presented below:

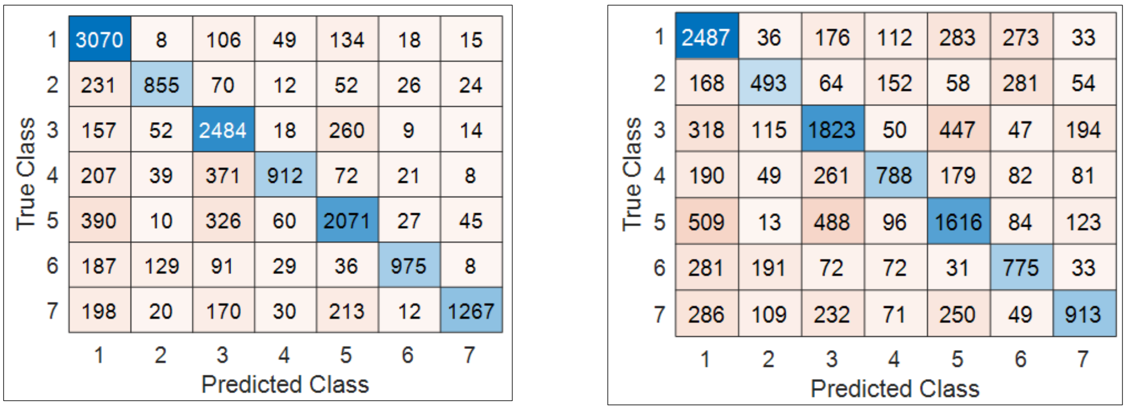
- Operating System: Windows Server 2019 Standard (64 bit)
- CPU: Intel(R) Xeon(R) CPU E5-2697 v2 @ 2.70 GHz (2 işlemci)
- RAM: 256 GB
- Platform: MATLAB 2021b

As mentioned in the other sections, 4 different prediction vectors (Chi2+SVM, Chi2+kNN, NCA+SVM and NCA+kNN) were obtained with the developed model. These prediction vectors were combined using the majority voting algorithm in the last phase and generalized classification results were obtained. At this stage, the confusion matrix of each vector was calculated to evaluate the performance of both the prediction vectors and the generalized classification results. In addition, various performance metric values were calculated using these matrices. These metrics are accuracy, recall, precision, F1 score and geometric mean.

In the test phase of the model, the 10-fold CV method was preferred as a validation strategy. The main reason for choosing this method is that all data is used in both training and test phases. The confusion matrices of 4 different prediction vectors obtained with the LBPStat-SelectFusion method developed in this context are given in Figure 3.



(a) Chi2+kNN (b) Chi2+SVM



(c) NCA+kNN **(d)** NCA+SVM

Figure 3 Confusion matrices obtained by combinations of feature selection and classification algorithms

With the developed model, four different classification results and accordingly four different prediction vectors are obtained. When the obtained confusion matrices are analyzed, it is seen that the highest classification accuracy is obtained in Figure 3-(a). This confusion matrix was obtained with the combination of Chi2+kNN algorithm. The performance metric values calculated using the confusion matrices obtained in this context are given in Table 2.

Table 3 Performance metric values

Case No	Combination	Metric	Value (%)
1	Chi2+kNN	Acc	75.34
		UAR	74.34
		UAP	76.01
		F1a	75.01
		Gm	73.89
2	Chi2+SVM	Acc	44.34
		UAR	39.22
		UAP	48.05
		F1a	40.60
		Gm	36.03
3	NCA+kNN	Acc	74.63
		UAR	71.51
		UAP	78.73
		F1a	73.90
		Gm	70.75
4	NCA+SVM	Acc	57.06
		UAR	53.92
		UAP	56.27
		F1a	54.64
		Gm	53.00

As seen in Table 2, the Chi2+kNN algorithm achieved the highest classification accuracy. On the contrary, the Chi2+SVM algorithm showed the lowest performance. In the last phase of the developed model, the majority voting algorithm was applied to increase the classification accuracy and generalize the results. This method takes four prediction vectors as input and produces the final classification result. In this context, the confusion matrix of the classification process obtained by applying majority voting is given in Figure 4.

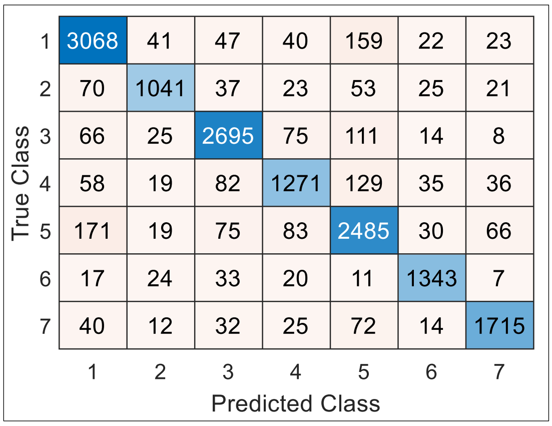


Figure 4 Confusion matrix of voted predicted vector

As shown in the figure, the performance of the classification process is further improved by majority voting. The performance metric values calculated using the confusion matrix given in Figure 4 are given in Table 3.

Table 4 Performance metric values for voted vector

Metric	Value (%)
Acc	87.36
UAR	86.73
UAP	87.54
F1a	87.11
Gm	86.59

As shown in Table 3, the classification accuracy of the developed model increased by >10% and reached >87% classification performance. This classification accuracy was achieved with the LBPStat-SelectFusion method, which has a very lightweight structure, and provided relatively low classification performance compared to the literature. However, the lightweight nature of the method and the fact that it does not involve any signal decomposition process keeps the temporal complexity of the developed model at a very low level and takes the model to the next level in terms of time/benefit. In this context, the prominent advantages, limitations and future work of the developed model are summarized below.

3.8.1. Findings and Advantages

- The LBPStat-SelectFusion method achieved 87.36% test accuracy on a dataset of real construction site equipment sounds.
- Due to the modular nature of the method, the ensemble-based (majority voting) classification strategy combined with different feature extraction and selection significantly improved the overall accuracy.
- The combination of LBP and statistical moment-based features strengthened the discrimination power between equipment types.
- The combined and parallel use of classical machine learning algorithms (kNN, SVM) has increased the computational efficiency and interpretability of the model.
- The method is practically applicable to large datasets with its low parameter count and fast training compared to deep learning architectures.

3.8.2. Limitations

- The dataset used was recorded in only one city (Elazığ) and under specific site conditions; the performance of the model should be further tested in different geographies or at construction sites with high ambient noise.
- Tested on only seven equipment types, accuracy may decrease with more equipment types or complex scene combinations.

3.8.3. Future Directions

- The method is planned to be tested on larger datasets collected at multiple sites, different cities and in various environmental conditions.
- Accuracy can be further improved by adding new time-frequency based features or automatic feature learning approaches to feature extraction.
- The model will be integrated into real-time monitoring systems and tested in industrial applications.
- Hybrid models can be developed with deep learning-based approaches.

3.8.4. Potential Implications

- Automated audio classification-based monitoring can significantly improve labor and resource optimization, occupational health and safety monitoring and process traceability in construction site management.
- The lightweight and easy-to-implement design offers usability on low-cost hardware, enabling widespread field applications.
- The outputs of the model can be used as a real-time decision support tool in future integrated construction site information systems and digital dashboards.

4. Conclusion

In this study, a novel and modular approach is developed to automatically classify equipment sounds recorded in a construction site environment using machine learning based methods. The proposed LBPStat-SelectFusion method is based on the processing of features extracted from both local binary patterns and statistical moments by two-way feature selection and fusion of multiple classifiers by majority voting. Experiments on an open-access dataset obtained under real construction site conditions show that the method performs well in equipment recognition tasks with high classification accuracy (87.36%).

In the field of construction management, digitalization of processes is becoming increasingly important in terms of increasing traceability and transparency. In this context, the integrated use of machine learning techniques with civil engineering and construction management significantly improves the effectiveness of project management by enabling automated, fast and reliable monitoring of field activities. The developed model provides a decision support mechanism that minimizes human error, optimizes resource utilization and identifies safety risks at an early stage compared to conventional methods based on manual observation.

As a result, the approach presented in this study sets an important example for the widespread use of machine learning-based automated equipment monitoring systems in construction management. In the future, the accuracy and generalizability of the model can be further strengthened with applications in different construction site environments and larger datasets. The combination of machine learning and construction management plays a critical role in achieving efficiency and sustainability goals in the construction industry.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] A. Pomaza-Ponomarenko, S. Kryvova, A. Hordieiev, A. Hanzyuk, and O. Halunko, "Innovative risk management: identification, assessment and management of risks in the context of innovative project management," 2023.

[2] M. Parsamehr, U. S. Perera, T. C. Dodanwala, P. Perera, and R. Ruparathna, "A review of construction management challenges and BIM-based solutions: perspectives from the schedule, cost, quality, and safety management," Asian Journal of Civil Engineering, vol. 24, no. 1, pp. 353-389, 2023.

[3] A. S. Rao et al., "Real-time monitoring of construction sites: Sensors, methods, and applications," Automation in Construction, vol. 136, p. 104099, 2022.

[4] V. Plevris, A. Ahmad, and N. D. Lagaros, Artificial intelligence and machine learning techniques for civil engineering. IGI Global, 2023.

[5] O. O. Abayomi-Alli, R. Damaševičius, A. Qazi, M. Adedoyin-Olowe, and S. Misra, "Data augmentation and deep learning methods in sound classification: A systematic review," Electronics, vol. 11, no. 22, p. 3795, 2022.

[6] J. Chukwunweike, A. N. Anang, A. A. Adeniran, and J. Dike, "Enhancing manufacturing efficiency and quality through automation and deep learning: addressing redundancy, defects, vibration analysis, and material strength optimization Vol. 23," World Journal of Advanced Research and Reviews. GSC Online Press, 2024.

[7] M. A. Musarat, A. M. Khan, W. S. Alaloul, N. Blas, and S. Ayub, "Automated monitoring innovations for efficient and safe construction practices," Results in Engineering, vol. 22, p. 102057, 2024.

[8] M. Kashef, A. Visvizi, and O. Troisi, "Smart city as a smart service system: Human-computer interaction and smart city surveillance systems," Computers in Human Behavior, vol. 124, p. 106923, 2021.

[9] O. Poudel and R. H. Assaad, "A Real-Time Intelligent Acoustic IoT-Enabled Embedded Construction Site Monitoring and Alert System: Integrating Deep Learning–Based Machine-Listening Algorithms, Edge Computing, and Cloud Computing," Journal of Construction Engineering and Management, vol. 151, no. 7, p. 04025075, 2025.

[10] N. R. Kapoor, A. Kumar, A. Kumar, A. Kumar, and H. C. Arora, "Artificial intelligence in civil engineering: An immersive view," in Artificial Intelligence Applications for Sustainable Construction: Elsevier, 2024, pp. 1-74.

[11] I. Ahmed, G. Jeon, and F. Piccialli, "From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where," IEEE Transactions on Industrial Informatics, vol. 18, no. 8, pp. 5031-5042, 2022.

[12] D. Ghai, S. L. Tripathi, S. Saxena, M. Chanda, and M. Alazab, Machine learning algorithms for signal and image processing. John Wiley & Sons, 2022.

[13] M. Scarpiniti, F. Colasante, S. Di Tanna, M. Ciancia, Y.-C. Lee, and A. Uncini, "Deep Belief Network based audio classification for construction sites monitoring," Expert Systems with Applications, vol. 177, p. 114839, 2021.

[14] A. Maccagno, A. Mastropietro, U. Mazziotta, M. Scarpiniti, Y.-C. Lee, and A. Uncini, "A CNN approach for audio classification in construction sites," in Progresses in Artificial Intelligence and Neural Systems: Springer, 2020, pp. 371-381.

[15] B. Sherafat, A. Rashidi, and S. Asgari, "Sound-based multiple-equipment activity recognition using convolutional neural networks," Automation in Construction, vol. 135, p. 104104, 2022.

[16] E. Akbal and T. Tuncer, "A learning model for automated construction site monitoring using ambient sounds," Automation in Construction, vol. 134, p. 104094, 2022.

[17] E. Akbal, P. D. Barua, S. Dogan, T. Tuncer, and U. R. Acharya, "DesPatNet25: Data encryption standard cipher model for accurate automated construction site monitoring with sound signals," Expert Systems with Applications, vol. 193, p. 116447, 2022.

[18] B. Sherafat, A. Rashidi, and S. Asgari, "Activity recognition of construction equipment using generated sound data," in Computing in Civil Engineering 2021, 2021, pp. 213-220.

[19] W. Xiong, X. Xu, L. Chen, and J. Yang, "Sound-based construction activity monitoring with deep learning," Buildings, vol. 12, no. 11, p. 1947, 2022.

[20] G. Jeong, C. R. Ahn, and M. Park, "Constructing an Audio Dataset of Construction Equipment from Online Sources for Audio-Based Recognition," 2022: IEEE, pp. 2354-2364.

[21] G. Jeong, M. Park, and C. R. Ahn, "Computer vision-aided audio dataset generation for recognizing construction equipment actions," Automation in Construction, vol. 171, p. 106014, 2025.

[22] M. Scarpiniti et al., "Leaky Echo State Network for Audio Classification in Construction Sites," in Applications of Artificial Intelligence and Neural Systems to Data Science: Springer, 2023, pp. 183-193.

[23] B. Nikmehr, M. R. Hosseini, I. Martek, E. K. Zavadskas, and J. Antucheviciene, "Digitalization as a strategic means of achieving sustainable efficiencies in construction management: A critical review," Sustainability, vol. 13, no. 9, p. 5040, 2021.

[24] N. Kaur and N. Nazir, "A review of local binary pattern based texture feature extraction," 2021: IEEE, pp. 1-4.

[25] M. Baygin, "An accurate automated schizophrenia detection using TQWT and statistical moment based feature extraction," Biomedical Signal Processing and Control, vol. 68, p. 102777, 2021.

[26] J. Goldberger, G. E. Hinton, S. Roweis, and R. R. Salakhutdinov, "Neighbourhood components analysis," Advances in neural information processing systems, vol. 17, 2004.

[27] H. Liu and R. Setiono, "Chi2: Feature selection and discretization of numeric attributes," 1995: Ieee, pp. 388-391.

[28] L. E. Peterson, "K-nearest neighbor," Scholarpedia, vol. 4, no. 2, p. 1883, 2009.

[29] W. S. Noble, "What is a support vector machine?," Nature biotechnology, vol. 24, no. 12, pp. 1565-1567, 2006.