Ecological Informatics 93 (2026) 103544

Contents lists available at ScienceDirect

Ecological Informatics

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

Predicting long-term environmental acoustic urban patterns using 2-slot short-term measurement and feed-forward artificial neural networks

Antonio Pita *, Juan M. Navarro

Escuela Politécnica Superior, UCAM Universidad Católica de Murcia, Av. de los Jerónimos, 135, Guadalupe, 30107, Murcia, Spain

ARTICLE INFO

Dataset link: https://opendata-ajuntament.barcelona.cat/data/ca/dataset/xarxasoroll-equipsmonitor-dades, https://github.com/AntonioPL/BCN_Noise

MSC:
68T07

Keywords:
Multivariate analysis
Machine learning
Environmental acoustics
Acoustic pattern prediction
Urban ecosystem monitoring

ABSTRACT

Monitoring acoustic environments in urban ecosystems poses a major challenge due to the temporal and spatial variability of soundscapes. Long-term data collection, often extending over a year, is recommended by regulations to establish reliable acoustic profiles, but such efforts are resource-intensive. In this study, we introduce a computational ecology approach to predict long-term acoustic patterns in cities using optimized combinations of time intervals as input for artificial neural networks. Unlike conventional methods relying on a single temporal window, our framework evaluates paired time intervals to enhance predictive performance and capture the dynamics of complex urban soundscapes. Multiple neural network architectures were designed and comparatively assessed, demonstrating that 2-slot datasets consistently improved classification accuracy and Balanced Accuracy Micro-Averaging across all categories. On average, temporal pairing increased Balanced Accuracy from 0.576 to 0.763 in the most variable category, reflecting a 32.4% improvement. These results highlight the importance of temporal diversity in ecological data modeling and underscore the potential of computational techniques to optimize temporary monitoring stations. The proposed method supports more efficient, data-driven strategies for environmental noise prediction, with direct implications for sustainable urban ecosystem management and decision-making in the context of global environmental change.

1. Introduction

Sound environments, or soundscapes, are increasingly recognized as a key ecological factor influencing the quality of life in urban ecosystems (Zipf et al., 2020). Excessive noise emissions represent one of the most pervasive environmental stressors, with implications not only for human well-being but also for biodiversity, species distribution, and the overall functioning of urban ecological systems. Urban acoustic environments present several major challenges: their strong spatial and temporal variability (Liu et al., 2013), the coexistence of multiple and overlapping sound sources, such as traffic, industrial, recreational, and natural ones, (Kang et al., 2016), and the need for detailed monitoring and predictive modeling (Aletta et al., 2016). Addressing these issues demands advanced data-driven strategies capable of capturing the complex dynamics of sound within urban ecosystems. Recent studies have highlighted the considerable spatio-temporal heterogeneity of soundscapes and the limitations of traditional monitoring networks, underscoring the importance of predictive and perception-aware approaches in environmental noise management (Basner and McGuire, 2018).

The European Directive 2002/49/EC establishes a framework for the standardized assessment and management of environmental noise, aiming to unify procedures and metrics across member states. Its primary objective is to mitigate, prevent, and reduce the adverse effects of noise pollution, including annoyance, for citizens exposed to various sound sources (European Commission, 2002). The directive encourages urban areas and city clusters to develop strategic noise maps (SNMs) and make the results publicly accessible. These maps serve as the foundation for action plans targeting noise reduction in areas with elevated noise exposure, known as noise exposure protection areas.

In recent years, many large cities have begun deploying wireless acoustic sensor networks (WASNs) using IoT technologies (Zanella et al., 2014). These networks are designed to collect sound data, which can be analyzed to refine SNMs and inform action plans for noise prevention and mitigation. WASNs typically consist of two main types of sensors: fixed-location stations for continuous, long-term monitoring and temporary stations for short-term assessments. Fixed stations are permanently installed at a single location, enabling continuous noise monitoring to detect trends and seasonal variations. In contrast, temporary stations are deployed at specific locations for predefined intervals (ranging from minutes to days) to capture short-term acoustic data. Temporary stations may include mobile sensors on vehicles equipped with geopositioning systems, standalone sensors deployed for limited

* Corresponding author.
E-mail address: apita@ucam.edu (A. Pita).

https://doi.org/10.1016/j.ecoinf.2025.103544
Received 22 September 2025; Received in revised form 29 November 2025; Accepted 29 November 2025
Available online 30 November 2025
1574-9541/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

durations, or conventional sound-measuring devices such as sound level meters (Garrido et al., 2019). However, using short-term data to identify environmental acoustic patterns is challenging because it often fails to account for seasonal effects, such as those associated with holidays or weekends (Fernandez-Prieto et al., 2020).

As recommended by Directive 2002/49/EC (European Commission, 2002), recognizing environmental acoustic patterns typically relies on long-term noise indicators, such as averaged or median values over at least a year. Based on these long-term statistics, two primary types of patterns are commonly identified: special regime areas, where noise levels exceed high thresholds, and quiet areas, where noise levels remain below low thresholds. In contrast, temporary stations are deployed to complement the spatial coverage of the fixed network by capturing short-term measurements at specific sites or periods; Their main purpose is to update and refine Strategic Noise Maps (SNMs) and to evaluate areas not permanently monitored. In this study, the information collected from temporary stations is further exploited through machine learning models to infer the long-term acoustic pattern of a site based on short-term observations, thereby enhancing the operational utility of temporary monitoring campaigns.

Previous studies have used supervised and unsupervised machine learning techniques to analyze sound in urban areas (Bianco et al., 2019), focusing primarily on classifying the acoustic environment (Bonet-Solà and Alsina-Pagès, 2021; Ullo et al., 2021). Moreover, machine learning methods have been widely applied to process audio data captured by stations to identify sources such as traffic, industrial activities, or natural sounds (Bilen et al., 2020; Nagy et al., 2020). Recent literature further highlights the versatility of machine learning in analyzing acoustic data for various purposes. For example, some studies focus on assessing traffic noise annoyance (Bravo-Moncayo et al., 2019), while others classify environmental noise (Albaji et al., 2023) or predict noise levels and emissions from household backup generators (Giwa et al., 2024). In urban contexts, machine learning has been effectively applied to monitor and predict traffic noise maps (Zambon et al., 2018) and analyze measurements to predict noise levels (Wen and Huang, 2020). Wireless sensor networks, combined with machine learning, have been used for long-term noise monitoring (Peckens et al., 2018) and acoustic event recognition (Luo et al., 2020), providing innovative approaches to tackle noise pollution challenges. In addition, unsupervised learning techniques have demonstrated potential in clustering WASN nodes based on similar acoustic behaviors (Pita et al., 2021, 2022). These clustering methods allow for the recognition of complex patterns, enabling the development of tailored strategies and the redefinition of acoustic zones in urban areas. Such advancements provide city managers with valuable tools to improve urban soundscapes and implement effective noise mitigation strategies.

In a previous study by the same authors (Navarro and Pita, 2023), the use of supervised machine learning techniques, specifically artificial neural networks (ANN), was proposed to predict the category of long-term acoustic behavior of a location based on short-term data. This approach enables temporary stations to estimate the long-term environmental acoustic pattern of a site, typically derived from year-long data, using only short-term acoustic measurements.

This proposed method accurately predicts the acoustic patterns of sites that showed low variability in sound pressure levels. However, a negative correlation was observed between variability and model performance, with patterns exhibiting higher variability resulting in lower prediction accuracy, which makes it difficult to identify some patterns of acoustic behavior.

The current research builds upon previous work with two main contributions: improving the overall prediction accuracy of the models and developing a systematic approach to predict behavioral patterns associated with high variability in sound pressure levels.

The current research builds upon previous work with two main contributions: improving the overall prediction accuracy of the models and developing a systematic approach to predict behavioral patterns associated with high variability in sound pressure levels. In this study, prediction accuracy is enhanced by introducing temporal diversity through the use of 2-slot short-term datasets, which combine non-contiguous hourly intervals to capture the daily variability of urban soundscapes. In addition, several feed-forward artificial neural network architectures with different depths and neuron configurations are evaluated to identify the most robust design for environmental acoustic pattern classification. Together, these contributions establish a systematic framework for modeling and predicting acoustic behaviors characterized by high variability in sound pressure levels, extending the capabilities of previous approaches based on single time-slot datasets.

2. Materials and methods

2.1. Data collection and processing

The data used in this study consists of sound pressure level measurements collected from 70 fixed acoustic nodes deployed across Barcelona (BCN), Spain, as part of a WASN. The spatial distribution of these nodes, covering representative urban functional areas such as residential, commercial, and leisure zones, was described and mapped in detail in a previous study (Pita et al., 2021), which also provides an interactive Python script for geographic exploration of node locations and spatial density. The same node configuration and spatial coverage were used in the present work to ensure methodological consistency. The design and deployment of this network are detailed in previous works by Camps (2015) and Farrés and Novas (2018). Each node is equipped with a Cesva TA120 sonometer (Anon, 2023), which continuously monitors sound pressure levels 24 h a day, 7 days a week. These devices comply with IEC 61672-1 Class 1 standards, providing an accuracy of ±1 dB and a resolution of 0.1 dB within a linear measurement range of 35 to 120 dBA. These devices calculate the A-weighted equivalent sound pressure level for each minute ($L_{\mathrm{Aeq1}m}$), as defined by ISO1996-2 (ISO 1996-2:2017, 2017). The formula for $L_{\mathrm{Aeq1}m}$ is expressed as:

$$L_{\mathrm{eq1}m} = 10 \cdot \log \left[\frac{1}{60} \int_{t_0}^{t_0+60} \frac{p^2(t)}{p_0^2} \mathrm{d}t \right] \ \mathrm{dBA}, \tag{1}$$

where $[t_0, t_0 + 60]$ represents a 1-minute interval starting at t_0, $p(t)$ is the sound pressure at time t in Pascals (Pa), and $p_0 = 20\ \mu Pa$ is the reference sound pressure level.

From 2018 to 2020, over 97 million $L_{\mathrm{Aeq1}m}$ records were captured by the 70 nodes and exported from the Barcelona smart city platform. The data were processed and standardized into consistent formats suitable for subsequent analysis, including several preprocessing steps to ensure quality and compatibility across nodes. During the curation phase, null values resulting from sensor failures and maintenance activities were removed to ensure data quality. In addition, a comprehensive exploratory data analysis (EDA) was conducted to assess the consistency and reliability of the measurements, using visual inspection through line plots, violin plots, and histograms to explore the distributions of sound pressure levels across sensors and time periods. To guarantee data reliability, measurements outside the operational limits specified by the Cesva TA120 manufacturer (below 35 dBA or above 120 dBA) were removed. Beyond these thresholds, no additional outlier filtering was applied, since in short-term acoustic captures it is generally not possible to identify outliers reliably. For instance, if construction activity occurs near a sensor, sound pressure levels may remain consistently elevated during the entire hour, making it impossible to distinguish them from valid but transient acoustic conditions. Nevertheless, during supervised short-term monitoring campaigns, operators can follow standardized measurement protocols to record contextual information and identify external factors, such as construction works, events, or adverse weather, that may affect the representativeness of the data. Further information on the basic statistics and record availability for each node is provided in Pita et al. (2021).

To train and evaluate the machine learning models, 300 curated datasets were prepared: 24 datasets representing one-hour time slots of the day (1-slot datasets), and 276 datasets representing all possible combinations of two distinct one-hour time slots (2-slot datasets). Each of the 24 1-slot dataset representing each hour, contains one instance/row for every combination of node × day during the study period, where each instance comprises 60 one-minute sound pressure level values corresponding to that hour as a columns. Similarly, each of the 276 2-slot dataset, representing a pair of hours, includes one instance for every node × day combination, with 120 one-minute measurements (60 for each hour) as a columns. For example, each instance in a 1-slot dataset for hour X (1-slot Xh) includes all sound pressure level values measured from X:00 to X:59 (hh:mm format), while each instance in a 2-slot dataset for hours X and Y (2-slot Xh & Yh) includes sound pressure level values from both X:00-X:59 and Y:00-Y:59 for all nodes and all days between January 2018 and December 2020.

To prepare these datasets for model training and evaluation using the machine learning technique outlined in Section 2.3, several preprocessing steps were applied. First, any hourly instances containing missing minute-level values were excluded to ensure that each dataset consisted only of complete one-hour observations. This step followed the initial curation phase, where individual minute-level null records were removed. Then, to preserve the temporal continuity of acoustic patterns, a chronological split was adopted instead of a random division. Specifically, all valid instances recorded during 2018 were used as the training set, while those from 2019 were reserved as the test set. This temporal partition ensures that the model is trained on past observations and evaluated on subsequent, non-overlapping data, thereby providing a more realistic assessment of predictive performance under real-world temporal conditions, as described in Section 2.4.

2.2. Environmental acoustic pattern

To assess the acoustic behavior of an environment, European Directive 2002/49/EC (European Commission, 2002) recommends using the L_{AeqT} indicators for the day, evening, and night periods over a 24-hour cycle, recorded daily for a full year at a specific monitoring station. These indicators are denoted as L_{d1y}, L_{e1y}, and L_{n1y}, respectively. Additionally, the directive introduces the composite indicator DEN (day, evening, night), represented as L_{den}, for an integrated assessment of noise levels over the specified period. To capture temporal variations in sound pressure levels across the daily intervals, the yearly standard deviation of L_{den1d}, denoted as $sd_{1y}(L_{den1d})$, has been proposed as a measure of annual variability or fluctuation in sound pressure levels (Pita et al., 2021, 2022). In these prior studies, these four acoustic indicators, calculated from the dataset, were used to identify several environmental acoustic patterns in Barcelona (BCN) and classify each acoustic node into one pattern category.

The behavioral categories of urban noise environments are classified based on sound pressure levels and variability. Category 1 includes high-traffic areas with steady noise levels and low variability, typically peaking during rush hours. Category 2 represents medium-density urban zones with moderate to high noise levels and variability, influenced by various activities. Category 3 corresponds to shopping and nightlife areas with consistently high and highly variable noise, driven by human and commercial activities. Finally, Category 4 refers to quiet residential zones with low noise levels and occasional variability due to transient sources. The average acoustic indicator values (centroids) that define these four behavioral categories are presented in Table 1.

Table 1 presents the mean values of the four noise indicators previously defined for each pattern category. The first category, consisting of 23 nodes, exhibits the highest sound pressure values (L_{d1y}, L_{e1y}, and L_{n1y}). The second category, which includes 27 nodes, shows intermediate sound pressure values, while the fourth category, comprising 11 nodes, has the lowest values. These categories thus correspond to nodes with high, medium, and low sound pressure levels, respectively, based on this consistent pattern. Additionally, a negative correlation is observed between sound pressure levels and their variability ($sd_{1y}(L_{den1d})$). The remaining third category, associated with nine nodes, shows a distinct behavior. In this group, evening sound pressure levels are higher than those during the day or night, which are relatively similar. Furthermore, this third category exhibits the greatest variability among all categories.

These environmental acoustic patterns can be related to urban characteristics such as road typology, land use, or the presence of specific noise sources. Identifying these relationships helps to interpret the meaning of each pattern within the urban context and to connect acoustic behaviors with the functional structure of the city. This type of analysis was conducted in a previous study (Pita et al., 2021), where the clusters were compared with the Strategic Noise Map (SNM) of Barcelona, demonstrating how the identified acoustic patterns correspond to real urban morphology and noise exposure zones. Understanding these relationships also highlights the importance of being able to predict acoustic behaviors, as it allows city managers to anticipate how changes in urban dynamics might affect environmental sound conditions.

2.3. Neural networks models

To estimate the target variable or behavior category described in Section 2.2, supervised learning algorithms were applied using the curated datasets outlined in Section 2.1. Specifically, multiple feed-forward multilayer artificial neural networks (McCulloch and Pitts, 1943) were developed and evaluated. In this study, the notation $Net\_X\_1\_\dots\_X\_n\_Y$ is used to represent a feed-forward ANN architecture with n hidden layers, where the ith hidden layer contains X_i neurons, and the output layer consists of Y neurons. Specifically, nine different architectures, with varying numbers of hidden layers and neuron configurations, were trained using the 300 datasets described in Section 2.1, resulting in a total of 2700 models for comparison. The nine architectures are as follows: Net_16_4, Net_32_4, Net_64_4, Net_16_16_4, Net_32_32_4, Net_64_32_4, Net_16_16_16_4, Net_32_32_32_4 and Net_64_32_16_4. It is important to note that all models share a common structure, with 120 neurons in the input layer and four neurons in the output layer.

Each hidden layer uses a Rectified Linear Unit (ReLU) activation function, while the output layer employs a Softmax activation function to produce class probabilities. The models were trained using the RMSprop optimizer with a learning rate of 0.001, momentum coefficient (ρ) of 0.9, and an epsilon value of 1×10^{-7}. The categorical cross-entropy loss function was used to minimize classification error, and accuracy was selected as the primary performance metric. These configurations were chosen to ensure numerical stability and efficient convergence across all trained models.

2.4. Performance metrics

To evaluate the performance of the models, several multiclass classification metrics were considered, providing a comprehensive assessment of their classification capabilities (Farhadpour et al., 2024; Grandini et al., 2008). These metrics include Accuracy, Balanced Accuracy Micro-Averaging, Precision, and Recall.

The metric notation is as follows: for each category i, TP_i represents the number of true positives, TN_i the true negatives, FP_i the false positives, and FN_i the false negatives for the binary classification problem of category i versus all. In addition, N denotes the total number of instances, and C represents the total number of categories.

Accuracy measures the proportion of correctly classified instances over the total number of instances. It is defined as:

$$\text{Accuracy} = \frac{\sum_{i=1}^{C} TP_i}{N} \tag{2}$$

Table 1
Environmental acoustic pattern classification of BCN WASN (adapted from Pita, Navarro, and Rodriguez Pita et al., 2021).

Category	L_{d2019}(dB)	L_{e2019}(dB)	L_{n2019}(dB)	$sd_{2019}(L_{den1d})$	Nodes
1	70.74	70.79	66.39	1.50	23
2	66.40	66.04	62.28	2.06	27
3	66.05	68.25	66.71	3.78	9
4	61.11	60.57	56.24	2.61	11

Precision quantifies the proportion of true positives among predicted positive instances, highlighting the ability of the model to avoid false positives. For each category i, it is defined as:

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \tag{3}$$

Recall, also known as sensitivity, measures the proportion of true positives correctly identified by the model among all actual positive instances. For each category i, it is defined as:

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \tag{4}$$

The Balanced Accuracy metric is used to evaluate the performance of binary classification models, particularly in the presence of class imbalance. It is calculated as the arithmetic mean of the True Positive Rate (TPR), also known as recall or sensitivity, and the True Negative Rate (TNR), also referred to as specificity or precision of the negative class. This approach ensures that the metric takes into account the classifier's ability to correctly predict both positive and negative classes. The formula for Balanced Accuracy, for each category i, is given by:

$$Balanced\,Accuracy_i = \frac{TPR_i + TNR_i}{2} \tag{5}$$

where:

$$TPR_i = \frac{TP_i}{TP_i + FN_i}$$

$$TNR_i = \frac{TN_i}{TN_i + FP_i}$$

Balanced Accuracy Micro-Averaging combines the concept of Balanced Accuracy, which accounts for class imbalance, with a micro-averaging approach that weights each class's contribution according to its size, i.e. treating all instances equally. It is calculated as a weighted average of the Balanced Accuracy for each class, with the weights corresponding to the number of instances in each class. The formula is given by:

$$\text{Balanced Accuracy Micro-Averaging} = \frac{\sum_{i=1}^{C} N_i \cdot \text{Balanced Accuracy}_i}{\sum_{i=1}^{C} N_i} \tag{6}$$

where C is the total number of classes, N_i is the number of instances in class i, $\text{Balanced Accuracy}_i = \frac{1}{2}\left(\frac{TP_i}{TP_i+FN_i} + \frac{TN_i}{TN_i+FP_i}\right)$ represents the Balanced Accuracy for class i, and TP_i, FN_i, FP_i, and TN_i are respectively the true positives, false negatives, false positives, and true negatives for class i. This formulation ensures that the contribution of each class to the overall metric reflects its representation in the dataset, making it particularly useful for evaluating models in datasets with imbalanced class distributions.

These metrics collectively provide a robust framework to assess the performance of the models, ensuring that both class-specific and overall evaluation criteria are addressed.

2.5. Evaluation

The evaluation process in this study is structured into three key stages, each aimed at assessing the datasets, comparing model performance, and identifying optimal conditions for prediction.

First, dataset evaluation is conducted by calculating the average accuracy of all models for each dataset. This analysis provides insight into dataset quality and its suitability for training predictive models. By examining the average accuracy across models, it is possible to determine whether 1-slot or 2-slot datasets are more effective for predicting environmental acoustic patterns. Second, model comparison focuses on identifying the architecture that achieves the highest performance across datasets. Accuracy serves as the primary evaluation metric, as it quantifies the proportion of correctly classified instances, offering an interpretable measure of model effectiveness. This analysis facilitates the selection of the most robust model for the classification task.

Table 2
Summary of performance metrics for each neural network model trained on 1-slot datasets and 2-slot datasets. The table includes the mean Accuracy and Balanced Accuracy Micro-Averaging for each model.

Model	Accuracy		BalAccuracyMA	
	1-slot	2-slot	1-slot	2-slot
Net_16_4	0.4127	0.4176	0.5470	0.5529
Net_32_4	0.3962	0.4028	0.5519	0.5576
Net_64_4	0.3669	0.3868	0.5264	0.5473
Net_16_16_4	0.4483	0.4592	0.5697	0.5840
Net_32_32_4	0.4589	0.4659	0.5815	0.5923
Net_64_32_4	0.4680	0.4720	0.5890	0.5948
Net_16_16_16_4	0.4751	0.5006	0.5989	0.6170
Net_32_32_32_4	0.4984	0.5084	0.6228	0.6268
Net_64_32_16_4	0.4581	0.4639	0.5699	0.5832

Finally, once the best-performing model is identified, further analysis is conducted using the Balanced Accuracy Micro-Averaging metric to determine the optimal time slot combinations. Balanced Accuracy Micro-Averaging accounts for class imbalances, providing a fairer assessment of model performance across different categories. Evaluating Balanced Accuracy Micro-Averaging for each time slot pair helps identify scenarios where the model achieves the highest predictive reliability, supporting more targeted applications.

This three-stage evaluation framework ensures a comprehensive analysis of both datasets and models, leading to a deeper understanding of the conditions under which the selected model performs optimally.

3. Results and discussion

3.1. Type dataset comparison

To evaluate dataset performance, models trained on 1-slot datasets were compared with those trained on 2-slot datasets. Table 2 summarizes the performance metrics for each model trained on both dataset types. For all nine ANN architectures, the models trained on 2-slot datasets outperformed their 1-slot counterparts in both Accuracy and Balanced Accuracy Micro-Averaging. The magnitude of this improvement varied across architectures. For instance, the largest increase was observed for the Net_64_4 model, rising from 0.3669 to 0.3868 (a relative improvement of 5.42%) in terms of Accuracy and from 0.5264 to 0.5473 (a relative improvement of 3.97%). A Wilcoxon signed-rank paired test was conducted to statistically validate this improvement, yielding a p-value of 0.001953 for the one-tailed comparison and 0.003906 for the two-tailed comparison. These results confirm that the 2-slot configuration systematically outperforms the 1-slot approach with statistical significance.

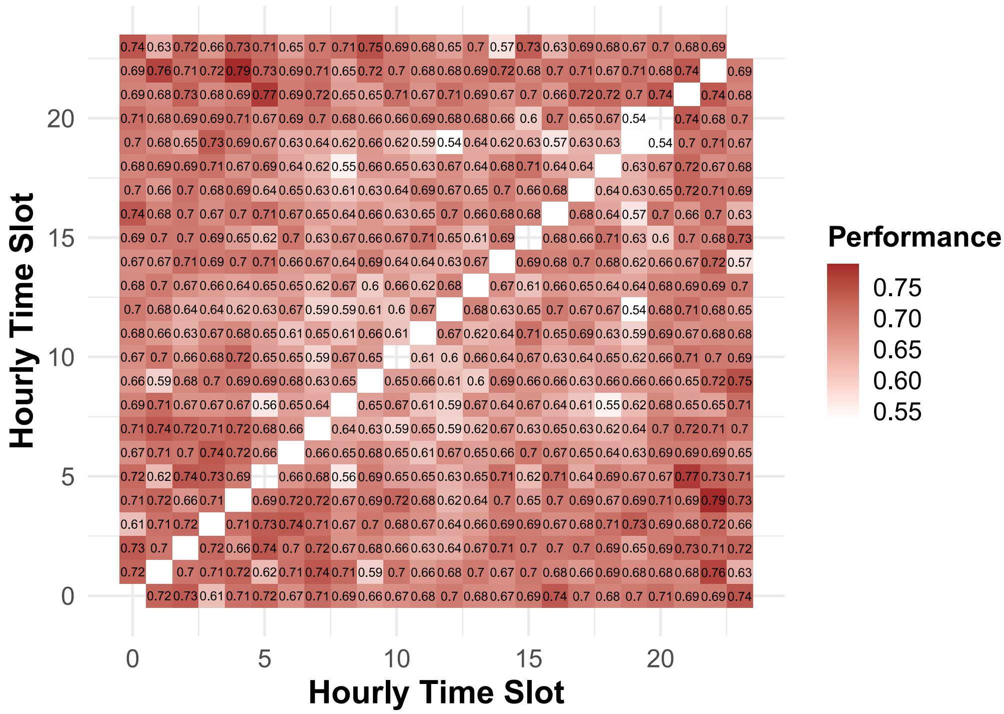

Fig. 1. heat-map showing the maximum Balanced Accuracy Micro-Averaging achieved by models for each 2-slot. Each cell represents a specific time slot pair, with the color intensity indicating the maximum Balanced Accuracy Micro-Averaging.

To further analyze model performance, the maximum Balanced Accuracy Micro-Averaging values obtained using the 2-slot datasets are presented in Fig. 1. The Balanced Accuracy Micro-Averaging metric, which accounts for class imbalance, exceeds 0.73 in the best-performing 2-slot configurations. The analysis indicates that the highest Balanced Accuracy Micro-Averaging values consistently occur for time slot pairs that include peak activity hours. Specifically, combinations such as 21 h–23 h (night) with 0 h–6 h (morning) yield the best results, with Balanced Accuracy Micro-Averaging values surpassing 0.73. These time periods likely capture a broader and more diverse range of acoustic patterns, improving the ability of the model to classify environmental acoustic behaviors accurately. In particular, Table 3 presents the top 15 configurations ranked by Balanced Accuracy Micro-Averaging. The highest recorded performance is 0.7868 for the 4 h and 22 h time slots, followed by the 5 h & 21 h and 01 h & 22 h configurations.

Conversely, the analysis also identified configurations in which model performance was considerably lower. The lowest Accuracy value among all 2-slot combinations was obtained with the `Net_32_4` architecture for the 16 h & 19 h pairing, reaching only 0.0942 and a Balanced Accuracy Micro-Averaging of 0.4980. These results highlight that temporal window selection is a critical factor, as certain hour pairs can substantially degrade model performance.

To assess the robustness of the obtained results and better understand the influence of the temporal configuration, a sensitivity analysis was conducted on the best-performing architecture (Net_32_32_32_4) (As will be shown in Section 3.2). For each pair of time slots, ten independent training runs were performed with randomized initializations, and the mean accuracy, standard deviation, and 95% confidence intervals were calculated. The results were then ordered by deciles to identify the most consistent temporal combinations. The analysis revealed that the highest-performing decile (Decile 10) was dominated by 2-slots combining nighttime and early-morning hours (1–5 a.m. and 20–23 p.m.). These pairs consistently achieved higher mean accuracies with lower standard deviations. For example, the 4 h & 22 h pairing achieved the highest mean Accuracy value of 0.6346, with a 95% confidence interval ranging from 0.6072 to 0.6619, whereas the 4 h & 8 h combination resulted in the lowest mean Accuracy (0.3548), with a substantially wider and lower confidence interval (0.2636–0.4460). These results indicate that the acoustic conditions during nighttime and early-morning periods are more stable and discriminative across urban categories, which facilitates the identification of long-term acoustic patterns from short-term measurements.

Table 3
Best 2-slot dataset by Balanced Accuracy Micro-Averaging Model Performance.

Model	2-slot	Accuracy	BalAccuracyMA
Net_32_32_32_4	04 h & 22 h	0.7018	0.7868
Net_16_16_16_4	05 h & 21 h	0.7012	0.7697
Net_16_16_16_4	01 h & 22 h	0.6921	0.7593
Net_32_32_32_4	09 h & 23 h	0.6606	0.7464
Net_32_32_16_4	05 h & 21 h	0.6644	0.7462
Net_32_16_8_4	00 h & 23 h	0.6646	0.7408
Net_32_32_32_4	01 h & 07 h	0.6464	0.7395
Net_16_16_16_4	02 h & 05 h	0.6469	0.7379
Net_32_32_16_4	03 h & 06 h	0.6384	0.7373
Net_64_32_16_4	00 h & 16 h	0.6476	0.7369
Net_32_32_32_4	21 h & 22 h	0.6640	0.7366
Net_16_16_16_4	20 h & 21 h	0.6603	0.7360
Net_32_32_32_4	00 h & 02 h	0.6408	0.7331
Net_32_32_32_4	03 h & 05 h	0.6431	0.7326
Net_64_32_16_4	04 h & 23 h	0.6511	0.7323

3.2. Neural network performance comparison

Finally, the performance of the nine neural network models trained on the 2-slot datasets is compared. These models are categorized based on their architecture: three with a single hidden layer (light blue), three with two hidden layers (blue), and three with three hidden layers (dark blue).

Fig. 2 shows a detailed visualization of the best-performing model for each 2-slot based on the highest Balanced Accuracy Micro-Averaging achieved. Each cell in the heat-map corresponds to a specific 2-slot,

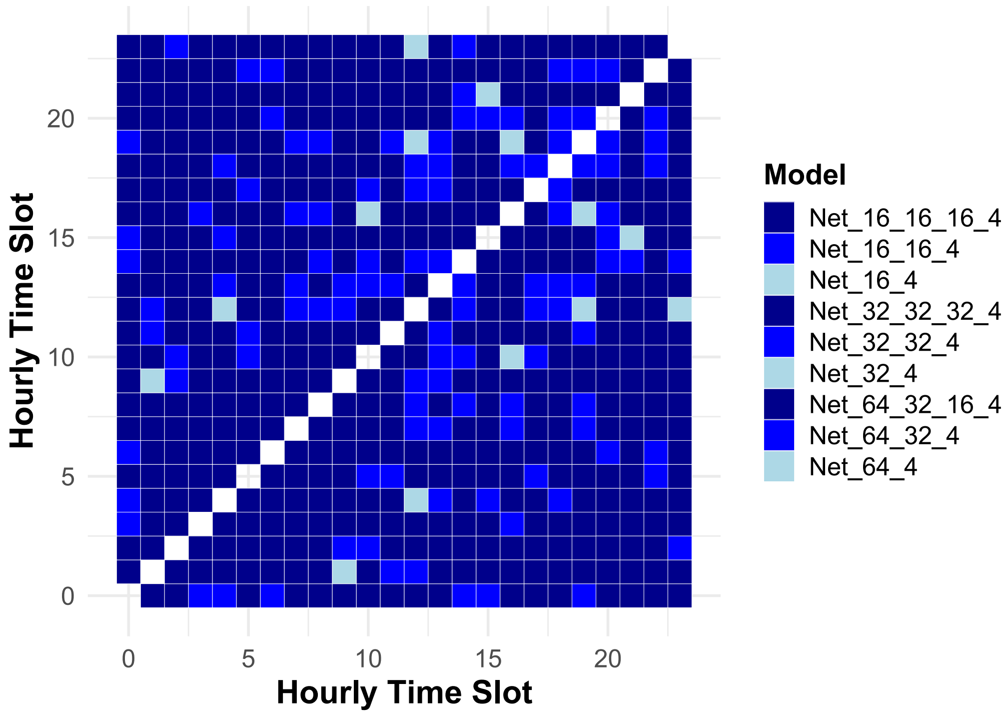

Fig. 2. Heat-map showing the best-performing neural network model for each 2-slot. Each cell represents a specific time slot pair, with colors indicating the model that achieved the highest Balanced Accuracy Micro-Averaging. Light blue denotes models with a single hidden layer, blue color models with two hidden layers, and dark blue color models with three hidden layers.

with colors representing the neural network architecture that attained the highest Balanced Accuracy Micro-Averaging. A gradient of blue shades differentiates the number of hidden layers: light blue denotes models with a single hidden layer, blue represents models with two hidden layers, and dark blue indicates models with three hidden layers.

This visualization enables a clear comparison of models performance across each 2-slot, demonstrating that the three-hidden-layer architecture outperforms others in most cases. Specifically, it achieves the highest Balanced Accuracy Micro-Averaging in 213 out of 276 2-slot datasets (77.17%), followed by the two-hidden-layer models, which perform best in 56 cases (20.29%), and the single-hidden-layer models, which achieve top performance in only 7 cases (2,54%). When considering the Accuracy metric, the three-hidden-layer architecture exhibits even greater dominance, leading in 216 out of 276 cases (78.26%). These results suggest that the additional complexity introduced by three hidden layers enhances the ability of the model to capture intricate patterns in the acoustic data. The consistent performance of this architecture highlights its robustness in modeling temporal relationships between paired time slots, further emphasizing the advantages of deeper neural networks for this task.

Fig. 3 summarizes the quantity of 2-slot datasets for which each model achieves the highest Balanced Accuracy Micro-Averaging. The $Net\_32\_32\_32\_4$ model stands out as the best-performing architecture, leading in 99 out of 276 2-slot datasets (35.87%). When considering the Accuracy metric, its dominance remains consistent, leading in 96 out of 276 cases (34.78%). These results highlight the reliability of the model and robustness across various temporal combinations, reinforcing its suitability for this classification task.

3.3. Category performance

Models trained on 1-slot datasets demonstrate inferior performance in predicting environmental behavioral categories characterized by higher variability in sound pressure levels (Navarro and Pita, 2023). This trend is particularly evident in Category 3, which corresponds to highly dynamic acoustic environments such as leisure or nightlife areas.

Table 4
Maximum Balanced Accuracy in each category on 1-slot datasets and 2-slot datasets.

Method	Category 1	Category 2	Category 3	Category 4
1-slot	0.8652	0.7536	0.5765	0.8601
2-slot	0.8835	0.7671	0.7631	0.8949

As shown in Table 4, the maximum Balanced Accuracy for Category 3 increased from 0.5765 (1-slot) to 0.7631 (2-slot), representing an improvement of approximately 32.4%.

This behavior is consistent with our previous findings (Navarro and Pita, 2023), where model performance was found to be inversely correlated with the daily variability of acoustic indicators, measured as $sd_{2019}(L_{den1d})$. Categories with higher volatility are inherently more difficult to predict due to their stochastic and transient nature. Incorporating a second time slot provides additional temporal context that enables the models to capture complex transitions and fluctuating sound dynamics more effectively, resulting in more stable and consistent predictions across variable urban sound environments. Future research could extend this analysis by examining the correlation between predictive improvement and variability indicators, as well as exploring adaptive temporal window strategies or confusion-pattern analyses to better understand inter-category boundaries and model misclassifications.

3.4. Computational performance

To complement the evaluation of predictive performance, the computational cost associated with training and deploying the proposed neural network architectures was also examined. On average, training a single model on a 1-slot dataset required 10.73 s, whereas training on a 2-slot dataset required 18.78 s. Although the individual training time per model is only moderately higher for 2-slot inputs, the total training time differs substantially because the number of 2-slot configurations

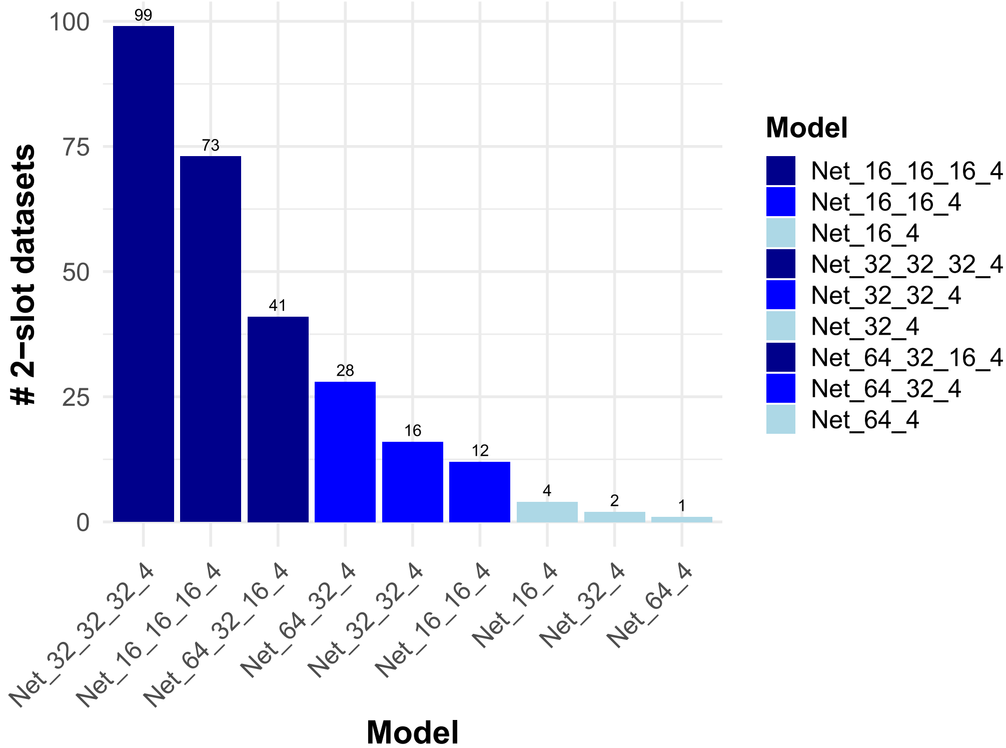

Fig. 3. Bar chart showing the quantity of 2-slot datasets where for which each neural network model achieves the highest Balanced Accuracy Micro-Averaging.

(276) is considerably larger than the number of 1-slot datasets (24). In total, training all 1-slot models (24 datasets × 9 architectures) required approximately 2317 s (38 min), while training all 2-slot models (276 datasets × 9 architectures) required 46649.52 s (777.49 min, or 12.95 h).

Despite this difference at dataset scale, the individual training cost remains low in both cases, and once a specific dataset and architecture are selected, the computational requirements do not pose a meaningful burden for practical applications. Furthermore, inference for all models operates on the order of milliseconds, confirming that the architectures are computationally lightweight and suitable for real-time or near-real-time deployment in urban acoustic monitoring systems. This efficiency is especially relevant for short-term mobile measurement campaigns, where rapid classification is necessary to assist field operators or automated monitoring workflows.

All computations were performed on an AMD Ryzen 7 5800X 8-Core CPU (3.80 GHz) with 32 GB RAM and an NVIDIA GTX 3050 GPU (8 GB GDDR6). The ANN models were implemented in Keras with a TensorFlow 2.x backend, using GPU acceleration for training. These hardware and software characteristics further indicate that the proposed methodology can be executed efficiently on widely accessible computing platforms, without requiring high-end infrastructure.

4. Conclusions

This study presented a methodology for predicting long-term environmental acoustic patterns using 2-slot short-term measurement and feed-forward artificial neural networks. The analysis assessed the effectiveness of various neural network architectures and dataset configurations to determine the most suitable approach for this classification task.

The results demonstrated that using 2-slot datasets to train and predict, consistently outperformed 1-slot datasets, both in average Accuracy and Balanced Accuracy Micro-Averaging metrics. This enhancement is especially pronounced in high-variability acoustic environments, such as those represented by Category 3, where the maximum Balanced Accuracy increased from 0.5765 to 0.7631 (32.4%). This finding aligns with prior evidence that highly volatile acoustic patterns are harder to predict and confirms that additional temporal context strengthens the model's ability to capture complex sound dynamics.

Among the nine ANN architectures evaluated, models with three hidden layers consistently outperformed those with one or two hidden layers, achieving the best performance in 213 out of 276 2-slot datasets (77.17%). Specifically, the $Net\_32\_32\_32\_4$ architecture demonstrated the highest overall effectiveness, emerging as the top-performing model in 99 out of 276 2-slot datasets (35.87%) based on Balanced Accuracy Micro-Averaging and in 96 out of 276 2-slot datasets (34.78%) based on Accuracy. Its robust design allowed it to fully utilize the temporal richness of the two-hour datasets, resulting in superior classification performance.

The analysis also revealed that the best-performing temporal pairs tend to involve nighttime and early-morning hours. These periods exhibit greater acoustic stability and stronger inter-category separability, which enhances model discriminability. A dedicated sensitivity analysis, based on repeated training runs, confidence intervals, and variability assessments, confirmed the stability of these results across random initializations and temporal splits. Specifically, the 2-slot combination of 4 h & 22 h resulted in a Balanced Accuracy Micro-Averaging of 0.7868 and an Accuracy of 0.7018.

In summary, the proposed approach effectively demonstrated the advantages of using 2-slot datasets and deep neural network architectures for environmental acoustic pattern recognition. The findings provide valuable insights for designing more efficient monitoring systems and support city managers in addressing noise pollution through enhanced temporal analysis.

The practical applicability of this methodology can be interpreted under two complementary scenarios. Cities that already operate wireless acoustic sensor networks and possess long-term data can use such information to train robust classification models and subsequently apply them with short-term ("2-slot") measurements to predict the acoustic behavior of sites without permanent sensors or to forecast future conditions, identifying potential behavioral changes. This aligns with the dynamic cluster evolution described in Cluster Analysis of Urban Acoustic Environments on Barcelona Sensor Network Data (Pita et al., 2021). Conversely, municipalities without long-term datasets may benefit from previously trained models, as urban areas have been shown to share recurrent acoustic behavior patterns (Pita et al., 2022). Such serialized models can be reused as baseline references, enabling short-term classification without waiting for a full year of data collection. In both contexts, the reuse of long-term data for initial training and the operational application of short-term measurements provide a scalable framework for cities to monitor, forecast, and manage their acoustic environments efficiently.

Taken together, these findings confirm that the proposed 2-slot methodology provides a practical and scalable approach for predicting long-term acoustic behavior from short-term measurements. While individual Balanced Accuracy Micro-Averaging values around 78% may not reflect perfect classification, their practical usefulness depends on the operational context. In real deployments, short-term mobile measurements can be repeated across different days and nearby locations, allowing temporal and spatial consistency checks that mitigate the impact of isolated misclassifications. This makes the approach a valuable complement to fixed monitoring networks, enabling cost-efficient large-scale noise assessment and the early detection of behavioral shifts in the urban soundscape.

Future research could extend the methodology presented in this study by examining whether the inclusion of three or more time slots continues to improve predictive performance or whether it reaches a plateau beyond a certain temporal resolution, thereby identifying an optimal balance between data granularity and model accuracy. Additionally, subsequent studies should assess the economic and computational implications of increasing temporal coverage, quantifying the trade-offs between accuracy, model complexity, and resource consumption in large-scale urban acoustic monitoring systems. Another relevant direction is the exploration of potential data imbalance issues, particularly when applying the methodology to cities or datasets with uneven class distributions. Incorporating balancing strategies — such as oversampling, class weighting, or synthetic data generation — may further enhance the robustness and generalization capability of the proposed framework.

CRediT authorship contribution statement

Antonio Pita: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Data curation, Conceptualization. **Juan M. Navarro:** Writing – review & editing, Writing – original draft, Validation, Supervision, Investigation.

Funding

This research was supported by ThinkInAzul project supported by Ministerio de Ciencia, Innovación y Universidades with funding from European Union NextGenerationEU under Grant PRTR-C17.I1 and from the Comunidad Autónoma de la Región de Murcia-Fundación Seneca.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors acknowledge Ajuntament of Barcelona for sharing the data.

Data and code availability

The data source described in this paper has been shared by the Ajuntament of Barcelona through their open data portal: https://opendata-ajuntament.barcelona.cat/data/ca/dataset/xarxasoroll-equipsmonitor-dades

The complete source code used for data processing, modeling, and analysis is available in the following GitHub repository: https://github.com/AntonioPL/BCN_Noise.

References

Albaji, O.A., Rashid, R., Abdul Hamid, S., 2023. Investigation on machine learning approaches for environmental noise classifications. J. Electr. Comput. Eng. 2023, 3615137. http://dx.doi.org/10.1155/2023/3615137.

Aletta, F., Kang, J., Axelsson, Ö., 2016. Soundscape descriptors and a conceptual framework for developing predictive soundscape models. Landsc. Urban Plan. 149, 65–74. http://dx.doi.org/10.1016/j.landurbplan.2016.02.001.

Anon, 2023. CESVA TA120 noise measuring sensor for smart solutions. Available online: https://www.cesva.com/en/products/sensors-terminals/TA120/. (Accessed 6 January 2023).

Basner, M., McGuire, S., 2018. WHO environmental noise guidelines for the European region: A systematic review on environmental noise and effects on sleep. Int. J. Env. Res. Public Health 15 (3), 519. http://dx.doi.org/10.3390/ijerph15030519.

Bianco, M.J., Gerstoft, P., Traer, J., Ozanich, E., Roch, M.A., Gannot, S., Deledalle, C.A., 2019. Machine learning in acoustics: Theory and applications. J. Acoust. Soc. Am. 146, 3590–3628. http://dx.doi.org/10.1121/1.5133944.

Bilen, Ç., Ferroni, G., Tuveri, F., Azcarreta, J., Krstulović, S., 2020. Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, Barcelona, Spain, pp. 61–65. http://dx.doi.org/10.48550/arXiv.1910.08440,

Bonet-Solà, D., Alsina-Pagès, R.M., 2021. A comparative survey of feature extraction and machine learning methods in diverse acoustic environments. Sensors 21, 1274. http://dx.doi.org/10.3390/s21041274.

Bravo-Moncayo, L., Lucio-Naranjo, J., Chávez, M., Pavón-García, I., Garzón, C., 2019. A machine learning approach for traffic-noise annoyance assessment. Appl. Acoust. 156, 262–270. http://dx.doi.org/10.1016/j.apacoust.2019.07.010.

Camps, J., 2015. Barcelona noise monitoring network. In: Proceedings of the EuroNoise. Maastricht, The Netherlands, pp. 218–220.

European Commission, 2002. Directive 2002/49/EC of the European Parliament and of the Council of 25 2002 Relating to the Assessment and Management of Environmental Noise. European Commission, Brussels, Belgium.

Farhadpour, S., Warner, T.A., Maxwell, A.E., 2024. Selecting and interpreting multiclass loss and accuracy assessment metrics for classifications with class imbalance: Guidance and best practices. Remote. Sens. 16, 533. http://dx.doi.org/10.3390/rs16030533.

Farrés, J.C., Novas, J.C., 2018. Issues and challenges to improve the barcelona noise monitoring network. In: Proceedings of the 11th European Congress and Exposition on Noise Control Engineering. Heraklion, Greece, pp. 27–31.

Fernandez-Prieto, J.A., Aguilar-García, G., Razavi, S., Leighton, T.G., Medina-Ramón, A., 2020. Wireless acoustic sensor nodes for noise monitoring in the built environment: Deployment considerations, data quality and seasonal trends. Sensors 20, 124. http://dx.doi.org/10.3390/s20010124.

Garrido, J.C., Mosquera, B.M., Echarte, J., Sanz, Roberto, 2019. Management noise network of madrid city council. In: InterNoise19, Proceedings of the Inter-Noise and Noise-Con Congress Conference, Madrid, Spain, 16–19 2019. Institute of Noise Control Engineering, Madrid, Spain, pp. 996–1997.

Giwa, S.O., Nwaokocha, C.N., Osifeko, O.M., Taziwa, R.T., Dyantyi, N., Sharifpur, M., 2024. Appraising machine learning algorithms in predicting noise level and emissions from gasoline-powered household backup generators. Int. J. Env. Sci. Technol. http://dx.doi.org/10.1007/s13762-024-05987-w.

Grandini, M., Bagli, E., Visani, G., 2008. Metrics for multi-class classification: An overview. http://dx.doi.org/10.48550/arXiv.2008.05756, arXiv preprint arXiv:2008.05756.

ISO 1996-2:2017, 2017. Acoustics—Description, Measurement and Assessment of Environmental Noise—Part 2: Determination of Environmental Noise Levels. International Organization for Standardization, Geneva, Switzerland.

Kang, J., Aletta, F., Gjestland, T.T., Brown, L.A., Botteldooren, D., Schulte-Fortkamp, B., Lercher, P., van Kamp, I., Genuit, K., Fiebig, A., Bento Coelho, J.L., Maffei, L., Lavia, L., 2016. Ten questions on the soundscapes of the built environment. Build. Env. 108, 284–294. http://dx.doi.org/10.1016/j.buildenv.2016.08.011.

Liu, J., Kang, J., Luo, T., Behm, H., Coppack, T., 2013. Spatiotemporal variability of soundscapes in a multiple functional urban area. Landsc. Urban Plan. 115, 1–9. http://dx.doi.org/10.1016/j.landurbplan.2013.03.008.

Luo, L., Qin, H., Song, X., Wang, M., Qiu, H., Zhou, Z., 2020. Wireless sensor networks for noise measurement and acoustic event recognitions in urban environments. Sensors 20, 2093. http://dx.doi.org/10.3390/s20072093.

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5, 115–133. http://dx.doi.org/10.1007/BF02478259.

Nagy, K., Cinkler, T., Simon, C., Vida, R., 2020. Internet of birds (iob): song based bird sensing via machine learning in the cloud: how to sense, identify, classify birds based on their songs?. In: Proceedings of the IEEE Sensors, Rotterdam, The Netherlands. pp. 1–4. http://dx.doi.org/10.1109/SENSORS47125.2020.9278714.

Navarro, J.M., Pita, A., 2023. Machine learning prediction of the long-term environmental acoustic pattern of a city location using short-term sound pressure level measurements. Appl. Sci. 13, 1613. http://dx.doi.org/10.3390/app13031613.

Peckens, C., Porter, C., Rink, T., 2018. Wireless sensor networks for long-term monitoring of urban noise. Sensors 18, 3161. http://dx.doi.org/10.3390/s18093161.

Pita, A., Rodriguez, F.J., Navarro, J.M., 2021. Cluster analysis of urban acoustic environments on Barcelona sensor network data. Int. J. Env. Res. Public Health 18, 8271. http://dx.doi.org/10.3390/ijerph18168271.

Pita, A., Rodriguez, F.J., Navarro, J.M., 2022. Analysis and evaluation of clustering techniques applied to wireless acoustics sensor network data. Appl. Sci. 12, 8550. http://dx.doi.org/10.3390/app12178550.

Ullo, S.L., Khare, S.K., Bajaj, V., Sinha, G.R., 2021. Hybrid computerized method for environmental sound classification. IEEE Access 8, 124055–124065. http://dx.doi.org/10.1109/ACCESS.2020.3006082.

Wen, P.-J., Huang, C., 2020. Noise prediction using machine learning with measurements analysis. Appl. Sci. 10, 6619. http://dx.doi.org/10.3390/app10186619.

Zambon, G., Roman, H.E., Smiraglia, M., Benocci, R., 2018. Monitoring and prediction of traffic noise in large urban areas. Appl. Sci. 8, 251. http://dx.doi.org/10.3390/app8020251.

Zanella, A., Bui, N., Castellani, A., Vangelista, L., Zorzi, M., 2014. Internet of things for smart cities. IEEE Internet Things J. 1, 22–32. http://dx.doi.org/10.1109/JIOT.2014.2306328.

Zipf, L., Primack, R.B., Rothendler, M., 2020. Citizen scientists and university students monitor noise pollution in cities and protected areas with smartphones. PLoS One 15, e0236785. http://dx.doi.org/10.1371/journal.pone.0236785.