A COLLABORATIVE SYSTEM FOR GEOREFERENCED MONITORING OF NOISE POLLUTION DATA CLASSIFIED BY MACHINE LEARNING MODEL

UM SISTEMA COLABORATIVO DE MONITORAMENTO GEORREFERENCIADO DE DADOS DE POLUIÇÃO SONORA CLASSIFICADOS POR MODELO DE APRENDIZADO DE MÁQUINA

UN SISTEMA COLABORATIVO DE MONITORIZACIÓN GEORREFERENCIADA DE DATOS DE CONTAMINACIÓN ACÚSTICA CLASIFICADOS POR MODELO DE APRENDIZAJE AUTOMÁTICO

https://doi.org/10.56238/sevened2026.008-158

Fani Tamires de Souza Batista[1], Sannyer Cardoso Carvalho Nery[2], Sérgio Augusto Coelho Bezerra[3], Keiko Verônica Ono Fonseca[4], José Reginaldo Hughes Carvalho[5]

ABSTRACT

This paper proposes a collaborative system for the georeferenced monitoring of noise pollution in urban environments, utilizing a machine learning model for automatic sound classification. The system adopts a modular architecture composed of three main applications: a Client Application (collection), a Server Application (processing), and a Client Application (visualization), which allows for scalability and ease of maintenance. In the collection Client Application, data capture is performed via a chatbot integrated with the Telegram application, enabling users to submit audio recordings along with their locations. These data are processed by a Server Application, which extracts acoustic features, classifies the sounds using the UrbanSound8K_ECAPA model, and stores the results in a NoSQL database. The visualization Client Application consists of a web page that requests data from the Server Application and presents the captured data interactively, allowing for the exploration of classified sounds and their corresponding geographical locations. A comparative analysis of audio classification models was conducted to support the choice of the most suitable model. Preliminary results demonstrate the viability of the proposed

---

[1] Graduated in Systems Analysis and Development Technology. Instituto Federal de Educação, Ciência e Tecnologia do Amazonas. E-mail: fani.tdsb@gmail.com
Orcid: https://orcid.org/0000-0002-0031-5040 Lattes: http://lattes.cnpq.br/9508646348735756

[2] Graduated in Systems Analysis and Development Technology. Instituto Federal de Educação, Ciência e Tecnologia do Amazonas. E-mail: carvalhosannyer@gmail.com Orcid: https://orcid.org/0009-0003-9169-5337
Lattes: http://lattes.cnpq.br/8514431967383778

[3] Doctoral student in Informatics. Universidade Federal de Minas Gerais (UFMG). Universidade Federal do Amazonas (UFAM). Instituto Federal de Educação, Ciência e Tecnologia do Amazonas.
E-mail: sergio.bezerra@ifam.edu.br Orcid: https://orcid.org/0000-0003-0887-6544
Lattes: http://lattes.cnpq.br/2717823426841852

[4] Dr. in Electrical Engineering. Universidade Federal de Santa Catarina (UFSC). Universidade Tecnológica Federal do Paraná. E-mail: keiko@utfpr.edu.br Orcid: https://orcid.org/0000-0003-4685-299X
Lattes: http://lattes.cnpq.br/8827013568855482

[5] Dr. in Electrical Engineering. Universidade Estadual de Campinas (UNICAMP). Universidade Federal do Amazonas. E-mail: reginaldo@icomp.ufam.edu.br Orcid: https://orcid.org/0000-0002-3828-7233
Lattes: http://lattes.cnpq.br/3161958119304780

solution, which proves to be a promising, low-cost, accessible, and scalable tool for mapping and analyzing urban noise pollution, with potential applications in public policy and smart city initiatives.

Keywords: Noise Pollution. Urban Sound Classification. Machine Learning. Collaborative Systems. Georeferencing.

RESUMO

Este trabalho propõe um sistema colaborativo para o monitoramento georreferenciado da poluição sonora em ambientes urbanos, utilizando um modelo de aprendizado de máquina para classificação automática de sons. O sistema adota uma arquitetura modular, composta por três aplicações principais: Aplicação Cliente (coleta); Aplicação Servidora (processamento) e Aplicação Cliente (visualização), o que permite escalabilidade e facilidade de manutenção. Na Aplicação cliente (coleta), a captura de dados é realizada por meio de um chatbot integrado ao aplicativo Telegram, permitindo que os usuários enviem gravações de áudio juntamente com suas localizações. Esses dados são processados por uma Aplicação Servidora, que extrai características acústicas e classifica os sons usando o modelo UrbanSound8K_ECAPA, e armazena os resultados em um banco de dados NoSQL. A Aplicação Cliente de visualização é composta por uma página Web, que faz requisições para a Aplicação Servidora e apresenta os dados capturados de forma interativa, permitindo a exploração dos sons classificados e suas correspondentes localizações geográficas. Uma análise comparativa de modelos de classificação de áudio foi conduzida para suportar a escolha do modelo mais adequado. Resultados preliminares demonstram a viabilidade da solução proposta, que se demonstra uma ferramenta promissora, de baixo custo, acessível e escalável para o mapeamento e análise da poluição sonora urbana, com potenciais aplicações em políticas públicas e iniciativas de cidades inteligentes.

Palavras-chave: Poluição Sonora. Classificação de Sons Urbanos. Aprendizado de Máquina. Sistemas Colaborativos. Georreferenciamento.

RESUMEN

Este trabajo propone un sistema colaborativo para el monitoreo georreferenciado de la contaminación acústica en entornos urbanos, utilizando un modelo de aprendizaje automático para la clasificación automática de sonidos. El sistema adopta una arquitectura modular, compuesta por tres aplicaciones principales: Aplicación Cliente (recolección); Aplicación Servidora (procesamiento) y Aplicación Cliente (visualización), lo que permite escalabilidad y facilidad de mantenimiento. En la Aplicación Cliente (recolección), la captura de datos se realiza a través de un chatbot integrado con la aplicación Telegram, permitiendo a los usuarios enviar grabaciones de audio junto con sus ubicaciones. Estos datos son procesados por una Aplicación Servidora, que extrae características acústicas y clasifica los sonidos usando el modelo UrbanSound8K_ECAPA, y almacena los resultados en una base de datos NoSQL. La Aplicación Cliente de visualización se compone de una página web, que realiza solicitudes a la Aplicación Servidora y presenta los datos capturados de forma interactiva, permitiendo la exploración de los sonidos clasificados y sus correspondientes ubicaciones geográficas. Se realizó un análisis comparativo de modelos de clasificación de audio para respaldar la elección del modelo más adecuado. Los resultados preliminares demuestran la viabilidad de la solución propuesta, que se perfila como una herramienta prometedora, de bajo costo, accesible y escalable para el mapeo y análisis de la contaminación acústica urbana, con potenciales aplicaciones en políticas públicas e iniciativas de ciudades inteligentes.

Palabras clave: Contaminación Acústica. Clasificación de Sonidos Urbanos. Aprendizaje Automático. Sistemas Colaborativos. Georreferenciación.

1 INTRODUCTION

Excessive noise is a form of noise pollution, which negatively affects the physical and mental health of the population. The high levels of decibels (dB), resulting from various human activities, generate noise that compromises the environmental silence and the sound balance of urban spaces (Silva; Cunha, 2022). The quantity that represents noise pollution is the Sound Pressure Level (SPL), measured in dB. And in Brazil, the ABNT NBR10151 (2019) establishes the SPL limits for urban ambient noise[6].

In view of these issues, in recent years, there has been a significant increase in the production of studies aimed at identifying noise levels in different contexts, such as work and residential environments and public roads. However, due to the high cost of physical devices used to measure noise pollution, the use of emerging technologies in this field has intensified.

In this scenario, the application of Artificial Intelligence (AI) techniques stands out, which have been consolidating themselves as effective tools for the analysis of large volumes of data in environmental monitoring systems (Santos *et al.*, 2020). The use of Machine Learning (ML)-based methods has high applicability in solving complex real-world problems, due to their potential to generate robust and accurate results (Beltran *et al.,* 2019).

This work proposes a system based on a machine learning model for classification and georeferenced monitoring of noise pollution data provided through mobile devices of collaborators in urban areas. The solution is based on a modular architecture consisting of three applications, allowing for separation of responsibilities and ease of maintenance.

2 RELATED JOBS

The works related to this research were organized into two categories, in order to support a better analysis of their proposals: (i) studies that do not incorporate classification models, that is, they monitor and/or map noise pollution through participatory systems (with or without the use of mobile devices); and (ii) studies that incorporate classification models for the monitoring of noise pollution, but that do not perform mapping or are not collaborative.

---

[6] For example, in a mixed, predominantly residential area, the limits are 55 dB(A) during the day and 50 dB(A) at night. The "A" index indicates that the measurement was performed using the "A" weighting, being an adjustment for measurements according to the perception of the human ear.

2.1 JOBS WITHOUT A CLASSIFICATION MODEL

By proposing a system for mapping traffic noise using *smartphones* as a low-cost alternative, Dubey *et al.* (2022) integrate the use of mobile devices and mapping techniques. In this context, the authors collected sound data at different times and locations in the city of Lucknow, India, performing the calibration of smartphones used with standard NPS meters, in order to improve the accuracy of the measurements. In other words, the calculation of noise levels is performed based on the recorded sound pressure compared to a reference level, shown in Equation 1 below:

$$NPS(dB) \;=\; 20 \cdot \; log_{\;10}\left(\frac{P}{P_{\;\;0}}\right) \tag{1}$$

Where:

SPL (dB) = Sound pressure level in decibels
P= Measured sound pressure
P0=Reference sound pressure

Thus, the authors demonstrated that the data obtained through smartphones, when properly calibrated, have variations of only 5 to 7 dB compared to conventional meters. Thus, the research concludes that smartphone-based solutions can be effective and affordable for mapping and monitoring urban noise pollution.

In the collaborative context, Silva and Cunha (2022) developed a project that proposes a low-cost solution for monitoring noise pollution in Belo Horizonte, using *crowdsourcing* through a Telegram chatbot. The proposal consists of sending voice messages and geolocation information, which are processed to compose a database on noise levels. The results allowed the mapping of the areas most affected by noise pollution and evidenced the accessibility of the solution, considering the minimum use of resources of the user's device.

In the study by Boumchich (2023), the use of smartphones as acoustic sensors for the evaluation of environmental noise is explored. The research focuses on the NoiseCapture 1 application, which allows users to measure and share noise data in real time. However, the quality and reliability of this data varies considerably. For this reason, the author proposes an analysis of the quality of the information by means of spatial grouping methods and by the implementation of an a posteriori calibration process, called "blind calibration".

In the work developed by Can (2023), a participatory methodology for evaluating the urban sound environment based on citizen action was presented. The process consisted of georeferenced noise measurement using smartphones, and the data collected were analyzed to create maps containing sound levels and sources. As a result, participatory acoustic maps were generated, which showed more accurately the diversity of urban sounds compared to traditional normative maps.

Still using mobile devices to capture sounds, Laranja and Saiter (2023) developed a study aimed at monitoring noise pollution in the city of Vitória, Espírito Santo, using smartphones with decibel meter applications. The results of the research demonstrated that noise levels often exceed the established limits, highlighting the importance of the elaboration of participatory noise maps for urban planning.

Finally, in the work of Sofianopoulos *et al.* (2024), the approach used was based on Volunteered Geographic Information (VGI) and web technologies to map noise pollution in Greece. Through the NoisePollutionApp application, citizens recorded sound intensity and their subjective perceptions about sounds, which made it possible to create collaborative maps.

2.2 WORK WITH CLASSIFICATION MODEL

Regarding the studies developed in recent years that monitor noise pollution with the incorporation of classification models, only three studies stand out.

In the research by Ahmed *et al.* (2021), a predictive model was proposed to estimate vehicular noise levels on Malaysian highways, using machine learning algorithms in conjunction with Geographic Information Systems (GIS). The models evaluated in this research were: Artificial Neural Network (ANN), Correlation-based Feature Selection with ANN (CFS-ANN) and the combination of Random Forest and ANN (RF-ANN). The authors considered factors such as vehicle type, traffic volume, temperature, humidity, topography, road density, and location. As a result, the Ensemble RF-ANN model stood out as the most accurate, showing that the combination of machine learning algorithms with GIS systems is an effective tool for predicting and mapping noise pollution in urban environments.

In the study developed by Ceriotti 2023, a system for classifying environmental sounds using convolutional neural networks (CNNs) was proposed, aimed at applications in embedded devices with limited computational resources. The author analyzed five different CNN architectures, applying different methods of extracting sound characteristics, with and without the use of *data augmentation* techniques. The database used was UrbanSound8K, composed of labeled recordings of urban sounds. Among the models tested, the Spatial Blended Convolutional Neural Network (SBCNN) was the one that presented the best performance.

Finally, in the project presented by Santos *et al.* (2023), the Intelligent Environmental Noise Monitoring (MIRA) platform was developed, consisting of sound monitoring stations, an integrated digital platform and artificial intelligence algorithms. The proposed solution allows the continuous recording of acoustic parameters from up to 100 monitoring stations, enabling the identification of patterns of variation in noise levels and their classification in almost real time.

3 AUDIO SIGNAL PROCESSING AND CLASSIFICATION

The processing and classification of audio signals involves advanced techniques of feature extraction and pattern modeling, often using the robustness of deep learning methods (Zaman *et al.*, 2023). In the context of the detection and classification of environmental sounds, the processing and classification of audio signals play a central role in the monitoring of phenomena such as noise pollution.

3.1 CLASSIFICATION-ORIENTED AUDIO PROCESSING

The processing of audio signals represents a fundamental step in sound classification systems, allowing the transformation of raw acoustic signals into numerical representations suitable for computational analysis. Thus, signal processing is a multidisciplinary field, combining principles from mathematics, statistics, computer science, and electrical engineering to extract relevant characteristics from acoustic signals (Arinaitwe *et al.,* 2024).

Audio processing for classification purposes usually follows a well-structured pipeline, consisting of essential steps such as acoustic signal acquisition, pre-processing, feature extraction, normalization, data *augmentation* techniques (where applicable), modeling, and finally, classification. Each of these phases plays a key role in building robust and accurate systems, and failures or deficiencies in any of them can significantly compromise the performance of the final model (Liu *et al.*, 2022; Mehrish *et al.*, 2023; Purwins *et al.*, 2019).

The feature extraction stage, for example, is a critical point in the process, as it is responsible for transforming the raw data of the sound signal into more compact and informative representations, which will serve as input for the learning algorithms. Techniques such as Mel-Frequency Cepstral Coefficients (MFCC), spectrograms, and learned embeddings have been widely employed in this context (Liu *et al.*, 2022; Purwins *et al.*, 2019).

In addition, the use of normalization and augmentation contributes to the generalization of models, allowing it to deal with variations in recording conditions and environmental noises. The proper definition and execution of this processing flow are therefore crucial for the success of automatic audio classification systems in real applications, such as environmental monitoring, sound event recognition, and virtual assistants (Mehrish *et al.,* 2023).

3.2 ANALYSIS OF SOUND CLASSIFICATION MODELS

The proposed system encompasses a classification model responsible for analyzing the audios collected by the client application. To verify the practical feasibility of the solution, this study considered the balance between effectiveness, efficiency and latency, as these criteria together indicate not only if the model is correct, but if it is operationally feasible for integration into the proposed system.

In this context, in order to verify the feasibility of incorporating Machine Learning techniques into the system, a comparative analysis of the main models for classifying environmental sounds present in the recent literature was carried out, evidencing their most relevant technical attributes. Thus, criteria such as the number of parameters, estimated latency, accuracy, number of classes, type of architecture, and the data set used in the validation experiments were considered.

The evaluation of the effectiveness of each model is carried out based on these criteria, with effectiveness being understood as the ability of an algorithm to produce accurate and reliable predictions, based on the comparison of performance metrics obtained through cross-validation (Sarker *et al.,* 2019). In this context, accuracy corresponds to the ratio between the number of correct predictions and the total number of predictions made (Strauss *et al.*, 2022).

Latency, in turn, refers to the average time required for a data packet to be transmitted from source to destination (Santos *et al.*, 2020). Efficiency, on the other hand, refers to the ability to optimally use the available resources and the existing structural conditions, aiming to achieve the best possible performance (Macedo *et al.,* 2010).

The selected models are widely used for the classification of urban sounds, being able to automatically identify and categorize different types of sounds captured in outdoor environments. Among the main models analyzed, the following stand out: UrbanSound8K_ECAPA, PANNs (Kong *et al.*, 2020), BEATs (Chen *at al.*, 2022), PaSST (Koutini *et al.,* 2022), AST (Gong *et al.*, 2021) and HTS-AT (Chen *at al.*, 2022).

Table 1

Technical characteristics of the analysis models

Model	Architecture	Dataset	Parameters
UrbanSound8k_ecapa	ECAPA-TDNN	UrbanSound8K	20.94 million
PANNs	CNN	AudioSet	80 million
BEATs (KS2)	Transformer	AudioSet	90 million
PaSST (ESC-50)	Transformer	ESC-50	87 million
AST	Transformer	AudioSet	31 million
HTS-AT (SpeechCmd V2)	Transformer	Speech Commands V2	31 million

Source: Prepared by the authors, 2025.

Table 1 presents the structure of these models aimed at the classification of environmental sounds and urban audio, focusing on the aspects of architecture, dataset and parameters used. The UrbanSound8K_ECAPA model uses an ECAPA-TDNN architecture, trained with the urban sounds of the UrbanSound8K dataset. The PANNs and PaSST models are based on the CNN and transformers family, respectively, using the AudioSet and the ESC-50, which are broader and more diverse.

AST and HTS-AT stand out for adapting originally visual architectures, such as *Vision Transformers* and *Swin Transformers*, to the acoustic domain, incorporating techniques such as the use of hierarchical patches. Finally, the BEATs model is the only one on the list to apply self-supervised learning, which makes it especially versatile in contexts with little labeled data. This diversity of structures and data evidences complementary approaches in the treatment of audio, ranging from specializations in urban noise to generalizations for multiple sound categories.

A relevant aspect illustrated in Table 1 concerns the number of parameters of the models, since this factor is directly related to structural complexity and its potential for generalization. The PANNs, BEATs, PaSST and AST models have a high number of parameters — ranging from 80 to 90 million — which implies high computational demand, too much training time, as well as high memory and energy consumption.

These characteristics can represent significant limitations for real-time applications or those that operate in environments with reduced computing infrastructure. On the other hand, the HTS-AT and UrbanSound8K_ECAPA models have between 20 and 31 million parameters, which makes them more viable options for systems with resource constraints, given their lower complexity.

Table 2 details the operational performance aspects of audio classification models. The observed accuracy shows that models such as HTS-AT (88.1%) and BEATs (87.6%) have superior performance in relation to most other approaches, standing out for the use of transformer-based architectures and unsupervised learning strategies. The PANNs model, on the other hand, although it has a deep architecture based on convolutional neural networks (CNN), has high latency, which can compromise its application in systems that require real-time responses (Kong *et al*, 2020).

Table 2

Performance of the models in the datasets used

Model	Latency	Accuracy (%)	Classes
UrbanSound8k_ecapa	Low to moderate	75,5	10
HTS-AT (SpeechCmd V2)	Low to moderate	97,2	35
PANNs	Moderate	89,5	527
BEATs (KS2)	Moderate	94,3	527
PaSST (ESC-50)	Moderate	93,2	527
AST	Moderate	95,6	527

Source: Prepared by the authors, 2025.

On the other hand, UrbanSound8K_ECAPA demonstrates efficiency even with a reduced number of classes, ten (Zhao and Liang, 2024), which makes it suitable for more restricted urban contexts and applications with a limited scope. The PaSST model has a good accuracy rate with moderate latency, configuring itself as a robust alternative, especially in scenarios with limited data availability (Koutini *et al.*, 2022). AST represents an innovative proposal by exploring the reuse of pre-trained weights in visual tasks for applications in the area of audio processing, obtaining promising results (Gong *et al.,* 2021).

In view of the above, the UrbanSound8K_ECAPA model was chosen in this research, based on three pillars: specificity, accessibility and efficiency. In terms of specificity, the UrbanSound8K database presents a detailed categorization of urban sounds commonly found in real environments, such as dog barking, children playing, air conditioning, street music, gunshots, car horns, sirens, gait motors, pneumatic hammers, and construction noises, which makes it suitable for the scope of the noise pollution study. This choice is directly linked to the need for accessible and computationally efficient models, aspects also observed in Macedo *et al.* (2024), which, as in this approach, adopt a perspective aimed at reducing computational complexity.

In this context, the selected model proves to be efficient, as it allows capturing complex acoustic patterns in audio classification tasks due to its robustness in extracting discriminative acoustic *embeddings* (Zhao and Liang, 2024). It is also worth mentioning that the accessibility to the UrbanSound8K_ECAPA model was a determining factor for its implementation in the proposed system.

4 METHODOLOGY

This work is a research of applied nature, since it seeks the production of knowledge and the elaboration of a collaborative system for monitoring georeferenced noise pollution. Regarding its objectives, the research is descriptive-exploratory, as it collects and classifies data on a phenomenon. In addition, it is concerned with identifying and evaluating collaborative systems that monitor noise pollution, as well as the search for recent studies in the area to map the problem and assist in the solution.

To carry out the search for related works, we opted for the adoption of a systematic process based on the works of Bezerra (2024, 2025), where the result of the selection of articles is represented by the PRISMA flowchart, adapted from Tricco *et al*. (2018), as shown in Figure 1. Initially, the research question was defined with the purpose of directing the entire process of identification and selection of studies. Next, the inclusion and exclusion criteria were established, considering aspects such as the period of publication, the availability of the full text, and the language in which the papers were published.

In the subsequent stage, the search strategies were outlined, which included the selection of specific databases, the definition of keywords and the application of Boolean operators, in order to ensure a comprehensive and effective search. Finally, the selection of studies was carried out, in which the studies initially identified were screened and later evaluated in their entirety.

Based on the delimitation of the research, the following guiding question was formulated: How can a collaborative system, capable of collecting, classifying and georeferencing urban sound data through mobile devices and machine learning techniques, contribute to the monitoring of noise pollution?

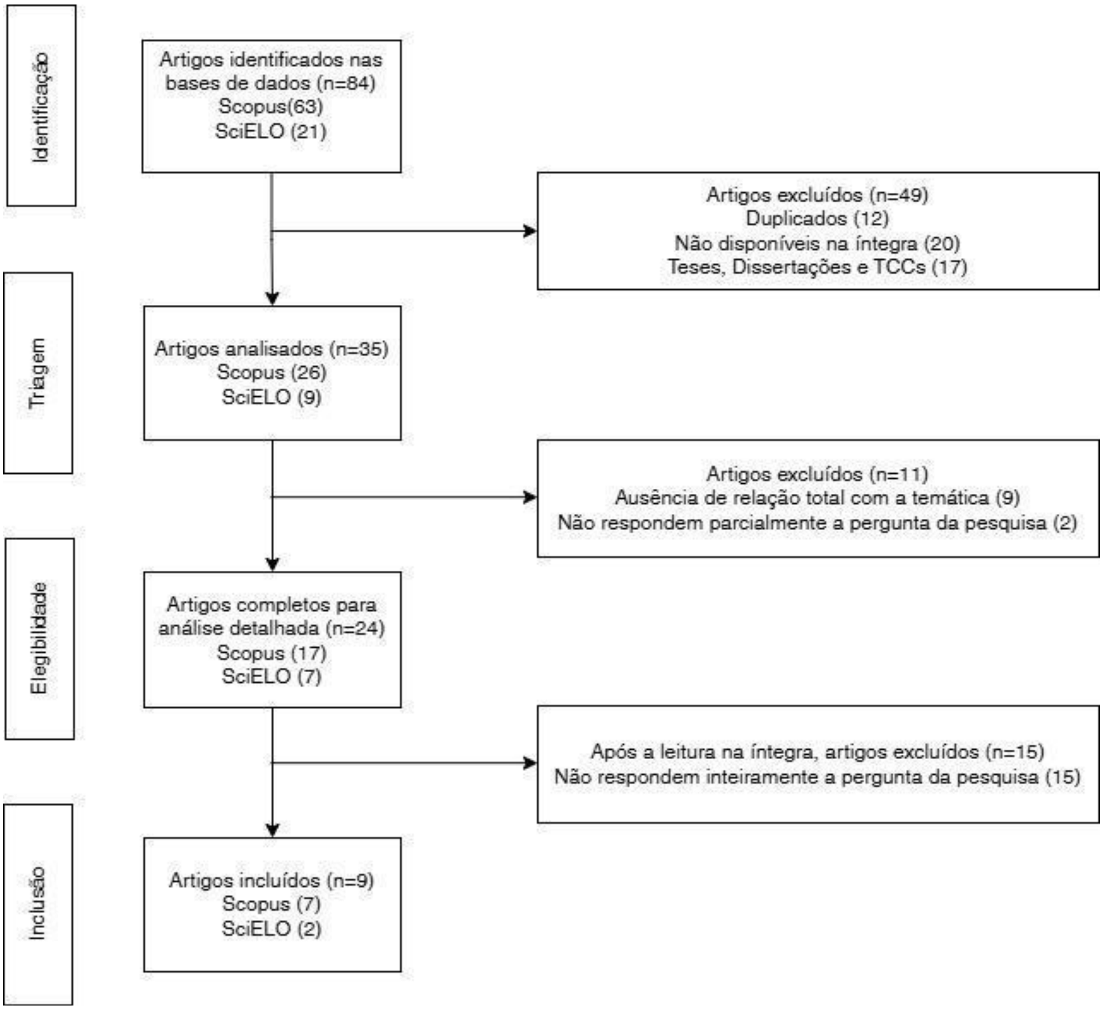
Based on this definition, the inclusion criteria were established, such as full-text studies, written in Portuguese or English, published between the years 2020 and 2024, which contained the descriptors defined in the title, abstract, or textual body. On the other hand, the exclusion criteria included duplicate, incomplete texts written in languages other than Portuguese and English, published before 2020, or that did not have the descriptors specified in any section of the study.

The search strategies involved conducting advanced searches in the Scopus and Scientific Electronic Library Online (Scielo) databases, using the following descriptors, in Portuguese and English: "System", "Mapping", "Machine Learning", "Machine Learning", "Deep Learning", "Deep Learning", "Collaborative", "Crowdsourcing", "Noise Pollution" and

"Noise Pollution". Thus, the analysis of the studies found was carried out through the reading of the abstracts, with the objective of verifying the relevance and adherence of the studies to the proposed theme.

Figure 1

Article selection flow



Source: Adapted from the 2020 PRISMA Guidelines and Bezerra *et al.* 2024.

5 PROPOSED SOLUTION

Compared to the solutions presented in the literature, the proposed system differs by modularly integrating the collection, processing and visualization of noise pollution data, using an audio classification model (UrbanSound8K) trained specifically for Brazilian urban environments. In addition, the present approach incorporates collaborative participation via Telegram chatbot, which broadens reach and reduces adoption barriers. Unlike systems that rely exclusively on dedicated sensors, the solution proposed here exploits the infrastructure of mobile devices already available, promoting low cost and greater scalability. Although no quantitative comparative evaluation was carried out at this stage, the preliminary analysis indicates that the proposal meets the same objectives of automatic classification, with the advantage of operating in an ecosystem that is easy to maintain and expand.

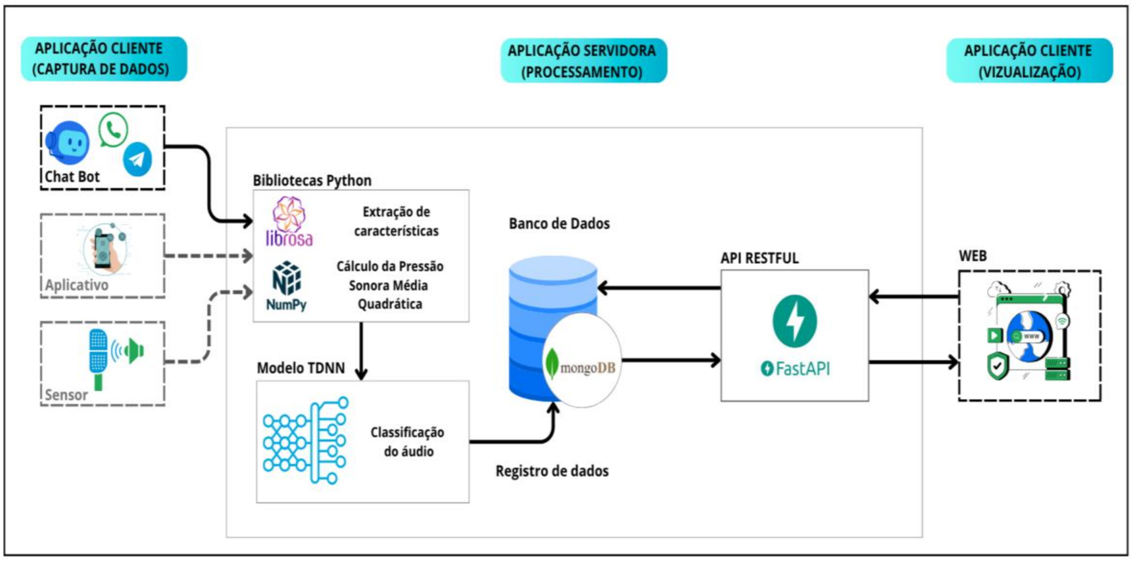
The proposed solution, as shown in Figure 2, presents the architecture of a system for monitoring noise pollution with georeferencing. The architecture is divided into three applications, with the objective of separating the responsibilities between data sending, processing and display of heat maps, namely: Client Application (Data Collection), Server Application (Processing) and Client Application (Visualization).

The Data Collection Client Application is responsible for capturing raw audio and location information. As illustrated in Figure 2, this collection can be carried out through chatbots integrated with messaging applications, such as WhatsApp and/or Telegram, dedicated applications or specific sensors. In this work, a chatbot was implemented on Telegram. It is worth mentioning that a chatbot offers a non-invasive form of data collection, since it does not require the installation of a new application. The user can start a conversation, record and send an audio containing the noise directly through the chat. Then, it is possible to share the location through the native functionalities of the messaging platform. This approach can impact the quality of sound data capture, since messaging apps are focused on recording human voice. In addition, the quality of smartphone microphones also influences data collection, as this hardware varies in constitution between brands and models. It is important to highlight that these impacts were not the object of studies in this work.

The Server Application is responsible for receiving the raw data, extracting relevant information, classifying the audios, storing them and making them available for consultation. For the processing of audio data, the Librosa library is used, a Python tool specialized in sound signal analysis. Among its functionalities, the extraction of acoustic characteristics, such as minimum and maximum sound pressure levels (SPL), in addition to the application of the "A" weighting filter, which simulates human auditory perception, stands out.

Figure 2

Architecture of the proposed solution



Source: Prepared by the authors, 2025.

In addition, the NumPy library, widely used for matrix and vector manipulation, is used in this work to calculate the Root Mean Square (RMS). This value represents the energy or "intensity" of the sound, and is later converted to the decibel scale (dB), the standard unit for noise measurement.

For the task of classifying the audios, the artificial intelligence model UrbanSound8K_ECAPA, based on Time Delay Neural Network (TDNN), is used. After extracting the acoustic characteristics, the audio files are converted from .ogg to .wav format, compatible with the adopted classification model. It is important to emphasize that a confidence rate of 60% was adopted as a criterion for validating the classifications.

The choice of this threshold represents a balance between accuracy, the ratio of correct ratings to positive ratings, and recall, the ability of the model to identify all relevant samples. A higher threshold would increase accuracy, but it could discard correct classifications with less confidence. The value of 60% was set to optimize the capture of relevant sound events without introducing an excessive number of false positives. Thus, when the classification reaches or exceeds this threshold, it is considered valid; otherwise, the audio is classified as belonging to the "Other" category.

After the extraction of the characteristics and the classification step, an object is created containing the information of the sending user, the geographic coordinates and the audio data. This object is then registered to an instance of MongoDB, a NoSQL database system that offers greater flexibility in storing and querying the data. Another component present in the Server Application, developed through the FastApi Framework according to the principles of Restful, is the Application Programming Interface (API), which acts as a communication bridge between stored data and external applications. This interface contains an endpoint that allows the Client Application (Preview) to access the processed data for display in the form of heat maps.

6 RESULTS AND DISCUSSIONS

The first component of the architecture, responsible for capturing data in the client application, was implemented as a chatbot on the Telegram platform. The system automatically registers users based on the information provided by the platform itself, also verifying that the user is already present in the database. In addition, the chatbot manages the sending of the audio samples along with their respective geographical locations.

Additionally, the choice of Telegram as a collection application presented advantages in terms of accessibility and user adherence, corroborating the approach adopted by Silva and Cunha (2022), who used Telegram's chatbot to collect environmental data in urban communities. This reinforces the potential of using technologies that are already consolidated for scientific purposes, without the need to develop native applications to test the operation of the application, which reduces costs and accelerates the prototyping cycle.

Figure 3

Registration of information in the Database

```
_id: ObjectId('6835b7a5642d9163184f8203')
id_user_object : ObjectId('6750565f07ed500dc84f821b')
latitude : -3.0407
longitude : -59.958332
data_criacao : "2025-05-27 09:00:48"
min_db : -80.87924194335938
max_db : -0.8792411088943481
min_dBA : -2.1170837671979967
max_dBA : 77.882916232802
rms_loudness : 0.0429314561188221
categoria : "children_playing"
```

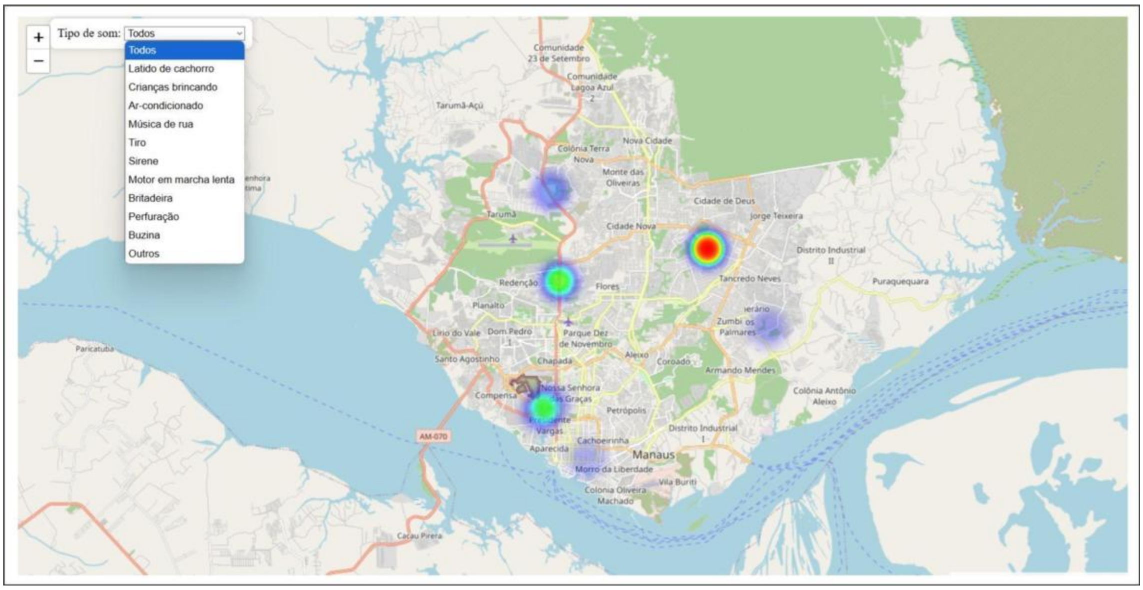
Source: Prepared by the authors, 2025.

The data processing proved to be effective, since the implementation was able to accurately extract (with a degree of conformity between the data entered) the essential information of each sample. This data includes the classification of the type of noise — performed through the machine learning model — geographic coordinates, the date and time of the shipment. Each consolidated record was stored in the MongoDB Atlas database, as Figure 3 illustrates.

With respect to visualization, the Client Application consumes the API and renders a dynamic heatmap, as evidenced in Figure 4. This application has a filtering feature by noise category, allowing a spatial analysis of noise pollution in a specific or general way of the type of sound.

Figure 4

Heat map



Source: Prepared by the authors, 2025.

In view of the above, this work shows the feasibility of the proposed system based on an architecture composed of three cohesive applications. This cohesion refers to the principles of microservices-based architecture, which favors independent and communicating services through APIs, promoting greater modularity, scalability, and resilience in distributed urban applications (Ribeiro, Braghetto, 2022). It is worth mentioning that the detailed operation of the applications that make up this system can be found in (Nery *et al.*, 2025).

7 CONCLUSION

This work consisted of the development of a collaborative system based on machine learning models for the classification and georeferenced monitoring of urban noise, received from mobile devices operated by volunteer collaborators.

Although the system has demonstrated efficacy in the correct classification of test audios, it is still necessary to carry out broader experiments, with a diversified data set, in order to improve the accuracy and generalization of the predictive model. In addition, the data visualization stage, although functional, will still undergo reformulations in order to improve the graphical interface. Such improvements aim to facilitate use by employees and public managers, expanding the social impact of the system. For future improvements in the accuracy of the collection, the development of a specialized application is indicated.

Thus, the potentiality of the applications developed, both on the client and server sides, indicates that the technology has applicability in various sectors of society. It is a low-cost, practical and scalable solution, whose implementation can help in the formulation of public policies, in the promotion of public health, in the support of scientific research and in environmental preservation, especially in regions affected by high levels of noise pollution.

ACKNOWLEDGMENTS

The authors would like to thank the Federal Institute of Education, Science and Technology of Amazonas (IFAM) and its Higher Education Course in Systems Analysis and Development (TADS). To the Graduate Program in Informatics (PPGI) of the Federal University of Amazonas and to the Graduate Program in Electrical Engineering and Industrial Informatics (CPGEI-CT) of the Federal Technological University of Paraná (UTFPR).

REFERENCES

Ahmed, A. A., Pradhan, B., Chakraborty, S., et al. (2021). Developing vehicular traffic noise prediction model through ensemble machine learning algorithms with GIS. Arabian Journal of Geosciences, 14(16), 1–15. https://doi.org/10.1007/s12517-021-08114-y

Arinaitwe, C., Mugisha, S., & Opio, D. (2024). Review of techniques used in speech signal processing. Journal of Signal Processing Systems, 3(1), 63–70. https://doi.org/10.59568/KJSET-2024-3-1-07

Beltran, C. A. R., et al. (2019). Plataforma de aprendizado de máquina para detecção e monitoramento de alunos com risco de evasão. Anais do CBIE/SBIE, 30(1), 1591. https://doi.org/10.5753/cbie.sbie.2019.1591

Bezerra, E. A. A. C., et al. (2025). Instrumentos construídos/validados ao processo de enfermagem em saúde mental: Revisão integrativa. Revista Eletrônica Acervo Saúde, 25, Article e19772. https://doi.org/10.25248/reas.e19772.2025

Bezerra, E. A. A. C., Ferreira, A. A., & Portela, L. C. (2024). Diagnóstico de enfermagem em saúde mental infantojuvenil. Miscellaneous, 24(1), Article e13700. https://doi.org/10.25248/reas.e13700.2024

Boumchich, A. (2023). A smartphone-based crowd-sourced database for environmental noise assessment: From data quality assessment to the production of relevant noise maps [Tese de doutorado, Le Mans Université]. https://theses.hal.science/tel-04444620/

Can, A., et al. (2023). Framework for urban sound assessment at the city scale based on citizen action, with the smartphone application NoiseCapture as a lever for participation. Noise Mapping, 10(1), Article 20220166. https://doi.org/10.1515/noise-2022-0166

Ceriotti, V. C. (2023). Classificação de sons ambientais utilizando redes neurais convolucionais para aplicações em hardwares com recursos limitados [Trabalho de

Conclusão de Curso, Universidade Federal do Rio Grande do Sul]. http://hdl.handle.net/10183/267582

Chen, K., Kong, Q., Iqbal, T., Wang, Y., Wang, W., & Plumbley, M. D. (2022). HTS-AT: A hierarchical token-semantic audio transformer for sound classification and detection. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 646–650). IEEE. https://doi.org/10.1109/ICASSP43922.2022.9746667

Chen, S., Wu, Y., Wang, C., Liu, S., Tompkins, D., Chen, Z., & Wei, F. (2022). BEATs: Audio pre-training with acoustic tokenizers [Preprint]. arXiv. https://arxiv.org/abs/2212.09058

Dubey, R., Bhardwaj, S., Sharma, V. B., Bhatt, A., & Biswas, S. (2022). Smartphone-based traffic noise mapping system. ISPRS Archives, XLIII-B4-2022, 613–620. https://doi.org/10.5194/isprs-archives-XLIII-B4-2022-613-2022

Gong, Y., Chung, Y.-A., & Glass, J. (2021). AST: Audio spectrogram transformer [Preprint]. arXiv. https://arxiv.org/abs/2104.01778

Kong, Q., Cao, Y., Iqbal, T., Wang, Y., Wang, W., & Plumbley, M. D. (2020). PANNs: Large-scale pretrained audio neural networks for audio pattern recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 2880–2894. https://doi.org/10.1109/TASLP.2020.3030497

Koutini, K., Schlüter, J., Eghbal-zadeh, H., & Widmer, G. (2022). Efficient training of audio transformers with patchout. In Interspeech 2022. ISCA. https://doi.org/10.21437/Interspeech.2022-227

Laranaja, A. C., & Saiter, J. C. (2023). Análise de poluição sonora urbana em Vitória–ES a partir do uso de smartphones. ENCAC, 17, 1–9. https://doi.org/10.46421/encac.v17i1.3885

Liu, S., Yuan, Y., Liu, X., & Liu, Y. (2022). Audio self-supervised learning: A survey. Patterns, 3(12), 1–18. https://doi.org/10.1016/j.patter.2022.100616

Macedo, A., Persona, L., & Meloni, F. (2024). Recognition of emotions through facial geometry with normalized landmarks. In Proceedings of the 30th Brazilian Symposium on Multimedia and the Web (pp. 257–266). Sociedade Brasileira de Computação. https://doi.org/10.5753/webmedia.2024.243252

Mehrish, A., Raj, R., Kumar, S., & Roy, P. P. (2023). A review of deep learning techniques for speech processing. Information Fusion, 99, 1–72. https://arxiv.org/abs/2305.00359

Nery, S. C. C., Batista, F. T. S., & Bezerra, S. A. C. (2025). ColabSomGeo: Ferramenta de monitoramento de poluição sonora subsidiada por modelo de aprendizagem. In Anais do 31º Simpósio Brasileiro de Sistemas Multimídia e Web. Sociedade Brasileira de Computação.

Purwins, H., et al. (2019). Deep learning for audio signal processing. IEEE Journal of Selected Topics in Signal Processing, 13(2), 206–219. https://doi.org/10.1109/JSTSP.2019.2908700

Ribeiro, M. B., & Braghetto, K. R. (2022). A scalable data integration architecture for smart cities: Implementation and evaluation. Journal of Information and Data Management, 13(2), 207–223. https://doi.org/10.5753/dm.2022.2485

Santos, A., Freitas, L. de, Teixeira, I., Vagner, G., Taira, G., Quile, R. E., & Bragheto, K. (2020). Desafios e oportunidades da aplicação de sistemas ciberfísicos no monitoramento da poluição urbana (pp. 276–289). https://doi.org/10.5753/courb.2020.12369

Santos, C. J., Souza, L. E., & Paiva, B. G. M. (2020). Análise da influência de diferentes codificadores de voz (CODECs) em redes convergentes. Revista Científica e-Locução, 1(17), 26–26. https://doi.org/10.57209/e-locucao.v1i17.249

Santos, L. C., Leão, C., Figueiredo, J., Melo, A., Covas, G., & Carvalho, A. (2023). Projeto MIRA – Desenvolvimento de uma plataforma inteligente de monitorização de ruído ambiental. https://www.dbwave.pt/uploads/media_items/mira-article-tecniacustica-20230915-lcs.original.pdf

Sarker, I. H., Kayes, A. S. M., & Watters, P. (2019). Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage. Journal of Big Data, 6, Article 57. https://doi.org/10.1186/s40537-019-0219-y

Silva, I. S. C., & Cunha, F. (2022). Monitoramento da poluição sonora em Belo Horizonte utilizando redes de sensoriamento participativo. https://bib.pucminas.br/pergamumweb/vinculos/00000b/00000bc6.pdf

Sofianopoulos, S., et al. (2024). Citizens as environmental sensors: Noise mapping and assessment on Lemnos Island, Greece, using VGI and web technologies. European Journal of Geography, 15(2), 106–119. https://www.eurogeojournal.eu/index.php/egj/article/view/476

Strauss, E., Villas Bôas Júnior, M., & Ferreira, W. L. L. (2022). A importância de utilizar métricas adequadas de avaliação de performance em modelos preditivos de machine learning. Projectus, 7(2), 52–62. https://doi.org/10.15202/25254146.2022v7n2p52

Tricco, A. C., et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Annals of Internal Medicine, 169(7), 467–473. https://doi.org/10.7326/M18-0850

Zaman, K., et al. (2023). A survey of audio classification using deep learning. IEEE Access, 11, 106620–106649. https://doi.org/10.1109/ACCESS.2023.3318015

Zhao, B., & Liang, J. (2024). Hierarchical-concatenate fusion TDNN for sound event classification. PLOS ONE, 19(10), 1–17. https://doi.org/10.1371/journal.pone.0312998