

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

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

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



10.56238/revgeov17n2-138

Fani Tamires de Souza Batista

Graduada em Tecnologia em Análise e Desenvolvimento de Sistemas
Instituição: 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>

Sannyer Cardoso Carvalho Nery

Graduado em Tecnologia em Análise e Desenvolvimento de Sistemas
Instituição: 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>

Sérgio Augusto Coelho Bezerra

Doutorando em Informática

Instituição: 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>

Keiko Verônica Ono Fonseca

Professora Doutora em Engenharia Elétrica

Instituição: 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>



José Reginaldo Hughes Carvalho

Professor Doutor em Engenharia Elétrica

Instituição: Universidade Estadual de Campinas (UNICAMP), Universidade Federal do Amazonas

E-mail: reginaldo@icomp.ufam.edu.brOrcid: <https://orcid.org/0000-0002-3828-7233>Lattes: <http://lattes.cnpq.br/3161958119304780>

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.

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 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.

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 INTRODUÇÃO

O excesso de ruídos se configura como uma forma de poluição sonora, a qual afeta negativamente a saúde física e mental da população. Os elevados níveis de decibéis (dB), resultantes de diversas atividades humanas, geram ruídos que comprometem o silêncio ambiental e o equilíbrio sonoro dos espaços urbanos (Silva; Cunha, 2022). A grandeza que representa a poluição sonora é o Nível de Pressão Sonora (NPS), medido em dB. E no Brasil, a NBR10151 da ABNT (2019) estabelece os limites de NPS para ruído ambiente urbano¹.

Diante dessas questões, observa-se, nos últimos anos, um aumento significativo na produção de estudos voltados à identificação dos níveis de ruído em diferentes contextos, como ambientes laborais, residenciais e vias públicas. Entretanto, devido ao elevado custo dos dispositivos físicos utilizados para a medição da poluição sonora, tem-se intensificado o uso de tecnologias emergentes nesse campo.

Nesse cenário, destaca-se a aplicação de técnicas de Inteligência Artificial (Artificial Intelligence, IA), que vêm se consolidando como ferramentas eficazes para a análise de grandes volumes de dados em sistemas de monitoramento ambiental (Santos *et al.*, 2020). O uso de métodos baseados em Aprendizado de Máquina (Machine Learning, ML) apresenta elevada aplicabilidade na resolução de problemas complexos do mundo real, em razão de seu potencial para gerar resultados robustos e precisos (Beltran *et al.*, 2019).

Este trabalho propõe um sistema baseado em modelo de aprendizado de máquina para classificação e monitoramento georreferenciado de dados de poluição sonora fornecido por meio de dispositivos móveis de colaboradores em áreas urbanas. A solução foi concebida com base em uma arquitetura modular composta por três aplicações, permitindo a separação de responsabilidades e facilidade de manutenção.

2 TRABALHOS RELACIONADOS

Os trabalhos relacionados com esta pesquisa foram organizados em duas categorias, a fim de subsidiar uma melhor análise quanto às suas propostas: (i) trabalhos que não incorporam modelos de classificação, ou seja, realizam o monitoramento e/ou o mapeamento da poluição sonora por meio de sistemas participativos (com ou sem o uso de dispositivos móveis); e (ii) trabalhos que incorporam modelos de classificação para o monitoramento da poluição sonora, mas que não realizam mapeamento ou não são colaborativos.

¹ Por exemplo, em uma área mista, predominantemente residencial, os limites são de 55 dB(A) durante o dia e 50 dB(A) à noite. O índice “A” indica que a medida foi realizada utilizando a ponderação “A”, sendo um ajuste para medidas segundo a percepção do ouvido humano.



2.1 TRABALHOS SEM MODELO DE CLASSIFICAÇÃO

Ao propor um sistema de mapeamento do ruído de tráfego por meio de *smartphones* como alternativa de baixo custo, Dubey *et al.* (2022) integram o uso de dispositivos móveis e técnicas de mapeamento. Nesse contexto, os autores coletaram dados sonoros em diferentes horários e locais da cidade de Lucknow, na Índia, realizando a calibração dos smartphones utilizados com medidores padrão NPS, a fim de melhorar a precisão das medições. Em outras palavras, o cálculo dos níveis de ruído é realizado com base na pressão sonora registrada em comparação a um nível de referência, demonstrado na Equação 1 a seguir:

$$NPS(dB) = 20 \cdot \log_{10} \left(\frac{P}{P_0} \right) \quad (1)$$

Onde:

NPS (dB) = Nível de pressão sonora em decibéis

P= Pressão sonora medida

P₀=Pressão sonora de referência

Dessa forma, os autores demonstraram que os dados obtidos por meio de smartphones, quando devidamente calibrados, apresentam variações de apenas 5 a 7 dB em comparação aos medidores convencionais. Assim, a pesquisa conclui que soluções baseadas em smartphones podem ser eficazes e acessíveis para o mapeamento e monitoramento da poluição sonora urbana.

No contexto colaborativo, Silva e Cunha (2022) desenvolveram um projeto que propõe uma solução de baixo custo para o monitoramento da poluição sonora em Belo Horizonte, utilizando *crowdsourcing* por meio de um chatbot do Telegram. A proposta consiste no envio de mensagens de voz e informações de geolocalização, as quais são processadas para compor um banco de dados sobre os níveis de ruído. Os resultados permitiram o mapeamento das áreas mais afetadas pela poluição sonora e evidenciaram a acessibilidade da solução, considerando o uso mínimo de recursos do dispositivo do usuário.

No estudo de Boumchich (2023), é explorado o uso de smartphones como sensores acústicos para a avaliação do ruído ambiental. A pesquisa concentra-se na aplicação NoiseCapture 1, que permite aos usuários medir e compartilhar dados de ruídos em tempo real. No entanto, a qualidade e a confiabilidade desses dados variam consideravelmente. Por esse motivo, o autor propõe uma análise da qualidade das informações por meio de métodos de agrupamento espacial e pela implementação de um processo de calibração a posteriori, denominado “calibração cega”.

No trabalho desenvolvido por Can (2023), foi apresentada uma metodologia participativa para avaliação do ambiente sonoro urbano com base na atuação cidadã. O processo consistiu na medição



georreferenciada dos ruídos por meio de smartphones, e os dados coletados foram analisados para a criação de mapas contendo níveis e fontes sonoras. Como resultado, foram gerados mapas acústicos participativos, os quais evidenciaram de forma mais precisa a diversidade dos sons urbanos em comparação aos mapas normativos tradicionais.

Ainda utilizando dispositivos móveis para a captação de sons, Laranja e Saiter (2023) desenvolveram um estudo voltado ao monitoramento da poluição sonora na cidade de Vitória, Espírito Santo, por meio de smartphones com aplicativos decibelímetros. Os resultados da pesquisa demonstraram que os níveis de ruído ultrapassam frequentemente os limites estabelecidos, ressaltando a importância da elaboração de mapas participativos de ruído para o planejamento urbano.

Por fim, no trabalho de Sofianopoulos *et al.* (2024), a abordagem utilizada baseou-se em Informações Geográficas Voluntárias (Volunteered Geographic Information, VGI) e em tecnologias Web para mapear a poluição sonora na Grécia. Por meio do aplicativo NoisePollutionApp, os cidadãos registravam a intensidade sonora e suas percepções subjetivas sobre os sons, o que possibilitou a criação de mapas colaborativos.

2.2 TRABALHOS COM MODELO DE CLASSIFICAÇÃO

No que se refere aos estudos desenvolvidos nos últimos anos que realizam o monitoramento da poluição sonora com a incorporação de modelos de classificação, destacam-se apenas três trabalhos.

Na pesquisa de Ahmed *et al.* (2021), foi proposto um modelo preditivo para estimar os níveis de ruído veicular em rodovias da Malásia, utilizando algoritmos de aprendizado de máquina em conjunto com Sistemas de Informação Geográfica (Geographic Information Systems, GIS). Os modelos avaliados nesta pesquisa foram: Artificial Neural Network (ANN), Correlation-based Feature Selection with ANN (CFS-ANN) e a combinação entre Random Forest e ANN (RF-ANN). Os autores consideraram fatores como tipo de veículo, volume de tráfego, temperatura, umidade, topografia, densidade das vias e localização. Como resultado, o modelo Ensemble RF-ANN destacou-se como o mais preciso, evidenciando que a combinação de algoritmos de aprendizado de máquina com sistemas GIS constitui uma ferramenta eficaz para a previsão e o mapeamento da poluição sonora em ambientes urbanos.

No estudo desenvolvido por Ceriotti 2023, foi proposto um sistema de classificação de sons ambientais com o uso de redes neurais convolucionais (Convolutional Neural Networks, CNNs), voltado para aplicações em dispositivos embarcados com recursos computacionais limitados. O autor analisou cinco arquiteturas distintas de CNNs, aplicando diferentes métodos de extração de características sonoras, com e sem a utilização de técnicas de *data augmentation*. A base de dados utilizada foi a UrbanSound8K, composta por gravações rotuladas de sons urbanos. Dentre os modelos testados, o Spatial Blended Convolutional Neural Network (SBCNN) foi o que apresentou melhor



desempenho.

Por fim, no projeto apresentado por Santos *et al.* (2023), foi desenvolvida a plataforma de Monitorização Inteligente de Ruído Ambiental (MIRA), composta por estações de monitoramento sonoro, uma plataforma digital integrada e algoritmos de inteligência artificial. A solução proposta permite o registro contínuo de parâmetros acústicos provenientes de até 100 estações de monitoramento, possibilitando a identificação de padrões de variação dos níveis de ruído e sua classificação quase em tempo real.

3 PROCESSAMENTO E CLASSIFICAÇÃO DE SINAIS DE ÁUDIO

O processamento e classificação de sinais de áudio envolve técnicas avançadas de extração de características e modelagem de padrões, utilizando frequentemente a robustez dos métodos de deep learning (Zaman *et al.*, 2023). No contexto da detecção e classificação de sons ambientais, o processamento e a classificação de sinais de áudio assumem papel central no monitoramento de fenômenos como a poluição sonora.

3.1 PROCESSAMENTO DE ÁUDIO VOLTADO À CLASSIFICAÇÃO

O processamento de sinais de áudio representa uma etapa fundamental em sistemas de classificação de som, permitindo a transformação de sinais acústicos brutos em representações numéricas adequadas para a análise computacional. Dessa forma, o processamento de sinais trata-se de um campo multidisciplinar, combinando princípios de matemática, estatística, ciências da computação e engenharia elétrica para extrair características relevantes dos sinais acústicos (Arinaitwe *et al.*, 2024).

O processamento de áudio para fins de classificação geralmente segue uma *pipeline* bem estruturada, composta por etapas essenciais como aquisição do sinal acústico, pré-processamento, extração de características, normalização, técnicas de data *augmentation* (quando aplicável), modelagem e, por fim, classificação. Cada uma dessas fases desempenha um papel fundamental na construção de sistemas robustos e precisos, sendo que falhas ou deficiências em qualquer uma delas podem comprometer significativamente o desempenho do modelo final (Liu *et al.*, 2022; Mehrish *et al.*, 2023; Purwins *et al.*, 2019).

A etapa de extração de características, por exemplo, constitui-se como um ponto crítico do processo, pois é responsável por transformar os dados brutos do sinal sonoro em representações mais compactas e informativas, que servirão como entrada para os algoritmos de aprendizado. Técnicas como Mel-Frequency Cepstral Coefficients (MFCC), espectrogramas e embeddings aprendidos têm sido amplamente empregadas nesse contexto (Liu *et al.*, 2022; Purwins *et al.*, 2019).

Além disso, o uso de normalização e augmentation contribui para a generalização dos modelos,



permitindo lidar com variações nas condições de gravação e ruídos ambientais. A definição e execução adequadas desse fluxo de processamento são, portanto, determinantes para o sucesso de sistemas automáticos de classificação de áudio em aplicações reais, como monitoramento ambiental, reconhecimento de eventos sonoros e assistentes virtuais (Mehrish *et al.*, 2023).

3.2 ANÁLISE DE MODELOS DE CLASSIFICAÇÃO DE SOM

O sistema proposto engloba um modelo de classificação responsável por analisar os áudios coletados pela aplicação cliente. Para verificar a viabilidade prática da solução, este estudo considerou o equilíbrio entre eficácia, eficiência e latência, pois esses critérios em conjunto, indicam não apenas se o modelo é correto, mas se é operacionalmente viável para integração ao sistema proposto.

Nesse contexto, com o objetivo de verificar a viabilidade da incorporação de técnicas de Aprendizado de Máquina ao sistema, realizou-se uma análise comparativa dos principais modelos de classificação de sons ambientais presentes na literatura recente, evidenciando seus atributos técnicos mais relevantes. Dessa forma, foram considerados critérios como: número de parâmetros, latência estimada, acurácia, número de classes, tipo de arquitetura e o conjunto de dados utilizado nos experimentos de validação.

A avaliação da eficácia de cada modelo é realizada com base nesses critérios, entendendo-se eficácia como a capacidade de um algoritmo produzir previsões precisas e confiáveis, a partir da comparação de métricas de desempenho obtidas por meio de validação cruzada (Sarker *et al.*, 2019). Nesse contexto, a acurácia corresponde à razão entre o número de previsões corretas e o número total de previsões realizadas (Strauss *et al.*, 2022).

A latência, por sua vez, diz respeito ao tempo médio necessário para que um pacote de dados seja transmitido de origem ao destino (Santos *et al.*, 2020). Já a eficiência refere-se à habilidade de empregar, de maneira otimizada, os recursos disponíveis e as condições estruturais existentes, visando alcançar o melhor desempenho possível (Macedo *et al.*, 2010).

Os modelos selecionados são amplamente utilizados para classificação de sons urbanos, sendo capazes de identificar e categorizar, de forma automática, diferentes tipos de sons captados em ambientes externos. Entre os principais modelos analisados, destacam-se: UrbanSound8K_ECAPA, PANNs (Kong *et al.*, 2020), BEATs (Chen *et al.*, 2022), PaSST (Koutini *et al.*, 2022), AST (Gong *et al.*, 2021) e HTS-AT (Chen *et al.*, 2022).



Tabela 1: Características técnicas dos modelos análise

Modelo	Arquitetura	Dataset	Parâmetros
UrbanSound8k_ecapa	ECAPA-TDNN	UrbanSound8K	20,94 milhões
PANNs	CNN	AudioSet	80 milhões
BEATs (KS2)	Transformer	AudioSet	90 milhões
PaSST (ESC-50)	Transformer	ESC-50	87 milhões
AST	Transformer	AudioSet	31 milhões
HTS-AT (SpeechCmd V2)	Transformer	Speech Commands V2	31 milhões

Fonte: Elaborado pelos autores, 2025.

A Tabela 1 apresenta a estrutura desses modelos voltados à classificação de sons ambientais e áudio urbano, com foco nos aspectos de arquitetura, dataset e parâmetros utilizados. O modelo UrbanSound8K_ECAPA utiliza uma arquitetura ECAPA-TDNN, treinado com os sons urbanos do dataset UrbanSound8K. Já os modelos PANNs e PaSST baseiam-se na família CNN e transformers, respectivamente, utilizando o AudioSet e o ESC-50, que são mais amplos e diversificados.

O AST e HTS-AT se destacam por adaptarem arquiteturas originalmente visuais, como *Vision Transformers* e *Swin Transformers*, para o domínio acústico, incorporando técnicas como o uso de patches hierárquicos. Por fim, o modelo BEATs é o único da lista a aplicar aprendizado auto-supervisionado, o que o torna especialmente versátil em contextos com poucos dados rotulados. Essa diversidade de estruturas e dados evidencia abordagens complementares no tratamento do áudio, indo desde especializações em ruídos urbanos até generalizações para múltiplas categorias sonoras.

Um aspecto relevante ilustrado na Tabela 1 diz respeito ao número de parâmetros dos modelos, uma vez que este fator está diretamente relacionado à complexidade estrutural e ao seu potencial de generalização. Os modelos PANNs, BEATs, PaSST e AST possuem um elevado número de parâmetros — variando entre 80 e 90 milhões — o que implica em alta demanda computacional, tempo de treinamento demasiado, bem como um alto consumo de memória e energia.

Essas características podem representar limitações significativas para aplicações em tempo real ou que operem em ambientes com infraestrutura computacional reduzida. Por outro lado, os modelos HTS-AT e UrbanSound8K_ECAPA apresentam entre 20 e 31 milhões de parâmetros, o que os torna opções mais viáveis para sistemas com restrições de recursos, dada sua menor complexidade.

A Tabela 2 detalha os aspectos de desempenho operacional dos modelos de classificação de áudio. A acurácia observada evidencia que modelos como HTS-AT (88,1%) e BEATs (87,6%) apresentam desempenho superior em relação à maioria das demais abordagens, destacando-se pelo uso de arquiteturas baseadas em transformers e estratégias de aprendizado não supervisionado. Já o modelo PANNs, embora apresente uma arquitetura profunda baseada em redes neurais convolucionais (CNN), possui alta latência, o que pode comprometer sua aplicação em sistemas que exigem respostas em tempo real (Kong *et al*, 2020).



Tabela 2: Desempenho dos modelos nos datasets utilizados.

Modelo	Latência	Acurácia (%)	Classes
UrbanSound8k_ecapa	Baixa a moderada	75,5	10
HTS-AT (SpeechCmd V2)	Baixa a moderada	97,2	35
PANNs	Moderada	89,5	527
BEATs (KS2)	Moderada	94,3	527
PaSST (ESC-50)	Moderada	93,2	527
AST	Moderada	95,6	527

Fonte: Elaborado pelos autores, 2025.

Em contrapartida, o UrbanSound8K_ECAPA demonstra eficiência mesmo com um número reduzido de classes, dez (Zhao e Liang, 2024), o que o torna adequado para contextos urbanos mais restritos e aplicações com escopo delimitado. O modelo PaSST apresenta uma boa taxa de acurácia com latência moderada, configurando-se como uma alternativa robusta, especialmente em cenários com disponibilidade limitada de dados (Koutini *et al.*, 2022). O AST representa uma proposta inovadora ao explorar o reaproveitamento de pesos pré-treinados em tarefas visuais para aplicações na área de processamento de áudio, obtendo resultados promissores (Gong *et al.*, 2021).

Diante do exposto, optou-se nesta pesquisa pelo modelo UrbanSound8K_ECAPA, com base em três pilares: especificidade, acessibilidade e eficiência. Em termos de especificidade, a base de dados UrbanSound8K apresenta uma categorização detalhada de sons urbanos comumente encontrados em ambientes reais, como latidos de cachorro, crianças brincando, ar-condicionado, música de rua, tiros, buzinas de carro, sirenes, motores de marcha, martelos pneumáticos e ruídos de construção, o que a torna adequada para o escopo do estudo de poluição sonora. Essa escolha está diretamente ligada à necessidade de modelos acessíveis e com eficiência computacional, aspectos também observados em Macedo *et al.* (2024), que, assim como nesta abordagem, adotam uma perspectiva voltada à redução de complexidade computacional.

Neste contexto, o modelo selecionado mostra-se eficiente, pois permite capturar padrões acústicos complexos em tarefas de classificação de áudio devido a sua robustez na extração de *embeddings* acústicos discriminativos (Zhao e Liang, 2024). Vale ressaltar, ainda, que a acessibilidade ao modelo UrbanSound8K_ECAPA foi fator determinante para sua implementação no sistema proposto.

4 METODOLOGIA

Este trabalho é uma pesquisa de natureza aplicada, uma vez que busca a produção de conhecimento e a elaboração de um sistema colaborativo de monitoramento da poluição sonora



georreferenciada. Em relação aos seus objetivos, a pesquisa é de caráter descritiva-exploratória, pois coleta e classifica os dados sobre um fenômeno. Além disso, preocupa-se em identificar e avaliar sistemas colaborativos que monitoram a poluição sonora, bem como a busca por estudos recentes na área para mapear o problema e auxiliar na solução.

Para a realização da busca por trabalhos relacionados, optou-se pela adoção de um processo sistemático baseado nos trabalhos de Bezerra (2024, 2025), onde o resultado da seleção dos artigos está representado pelo fluxograma PRISMA, adaptado de Tricco *et al.* (2018), conforme a Figura 1. Inicialmente, foi definida a pergunta da pesquisa com o propósito de direcionar todo o processo de identificação e seleção dos estudos. Em seguida, foram estabelecidos os critérios de inclusão e exclusão, considerando aspectos como o período de publicação, a disponibilidade do texto completo e o idioma em que os trabalhos foram publicados.

Na etapa subsequente, foram delineadas as estratégias de busca, que incluíram a seleção de bases de dados específicas, a definição de palavras-chave e a aplicação de operadores booleanos, com o intuito de assegurar uma busca abrangente e eficaz. Por fim, realizou-se a seleção dos trabalhos, na qual os estudos inicialmente identificados foram triados e, posteriormente, avaliados em sua totalidade.

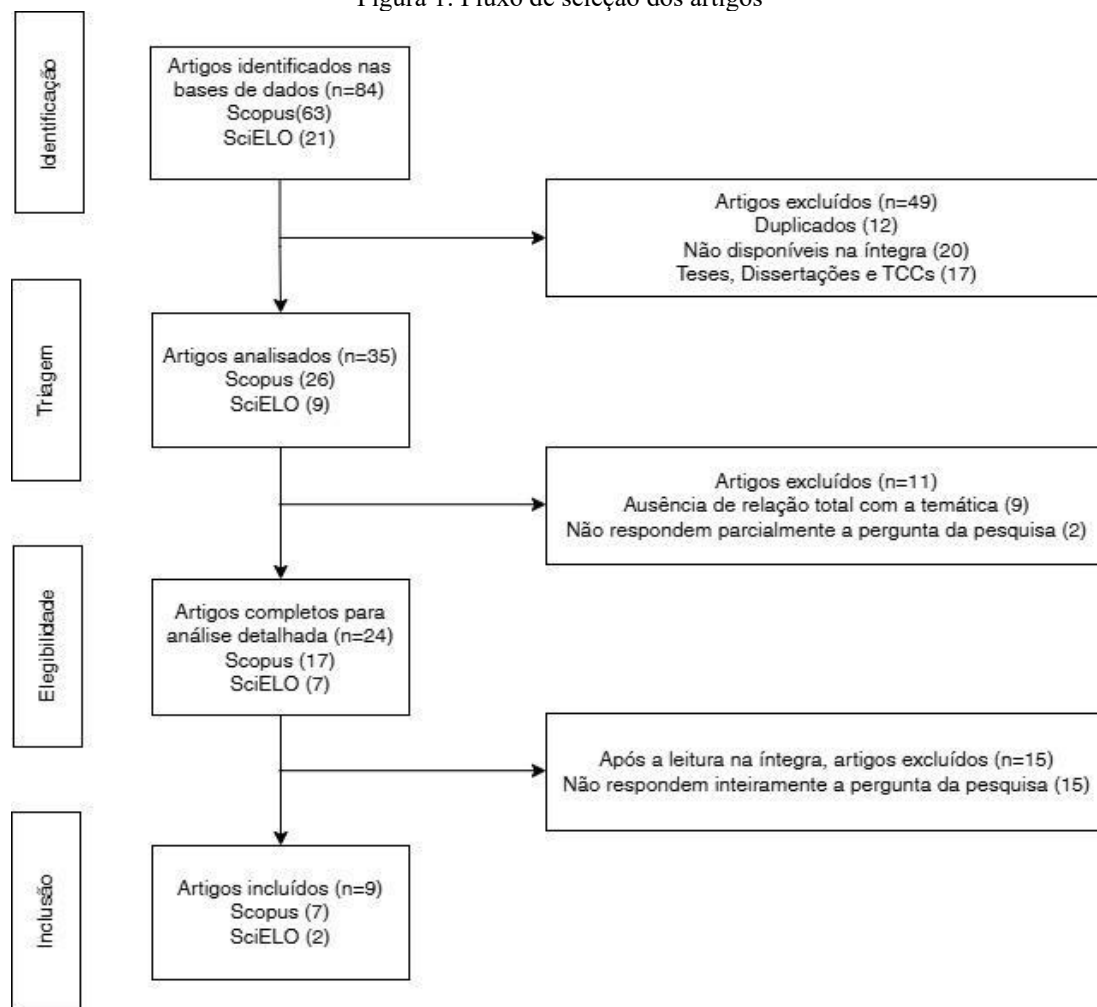
Com base na delimitação da pesquisa, formulou-se a seguinte pergunta norteadora: Como um sistema colaborativo, capaz de coletar, classificar e georreferenciar dados sonoros urbanos por meio de dispositivos móveis e técnicas de aprendizado de máquina, pode contribuir para o monitoramento da poluição sonora?

A partir dessa definição, foram estabelecidos os critérios de inclusão, como estudos com texto completo, redigidos em português ou inglês, publicados entre os anos de 2020 e 2024, que contivessem os descritores definidos no título, resumo ou corpo textual. Por outro lado, os critérios de exclusão abrangeram textos duplicados, incompletos, redigidos em idiomas distintos do português e do inglês, publicados antes de 2020 ou que não apresentassem os descritores especificados em nenhuma seção do trabalho.

As estratégias de busca envolveram a realização de pesquisas avançadas nas bases de dados Scopus e Scientific Electronic Library Online (SciELO), utilizando os seguintes descritores, em português e inglês: “Sistema”, “System”, “Mapeamento”, “Mapping”, “Aprendizado de Máquina”, “Machine Learning”, “Aprendizado Profundo”, “Deep Learning”, “Colaborativo”, “Crowdsourcing”, “Poluição Sonora” e “Noise Pollution”. Assim, a análise dos trabalhos encontrados foi realizada por meio da leitura dos resumos, com o objetivo de verificar a relevância e aderência dos estudos à temática proposta.



Figura 1: Fluxo de seleção dos artigos



Fonte: Adaptado das Diretrizes de PRISMA 2020 e Bezerra *et al.* 2024.

5 SOLUÇÃO PROPOSTA

Em comparação com as soluções apresentadas na literatura, o sistema proposto diferencia-se por integrar de forma modular a coleta, processamento e visualização de dados de poluição sonora, utilizando um modelo de classificação de áudio (UrbanSound8K) treinado especificamente para ambientes urbanos brasileiros. Além disso, a presente abordagem incorpora participação colaborativa via chatbot do Telegram, o que amplia o alcance e reduz barreiras de adoção. Diferentemente de sistemas que dependem exclusivamente de sensores dedicados, a solução aqui proposta explora a infraestrutura de dispositivos móveis já disponíveis, promovendo o baixo custo e maior escalabilidade. Embora não tenha sido realizada avaliação quantitativa comparativa nesta etapa, a análise preliminar indica que a proposta atende aos mesmos objetivos de classificação automática, com a vantagem de operar em um ecossistema de fácil manutenção e expansão.

A solução proposta, como mostra a Figura 2, apresenta a arquitetura de um sistema para monitoramento da poluição sonora com georreferenciamento. A arquitetura está dividida em três aplicações, com o objetivo de separar as responsabilidades entre envio de dados, processamento e

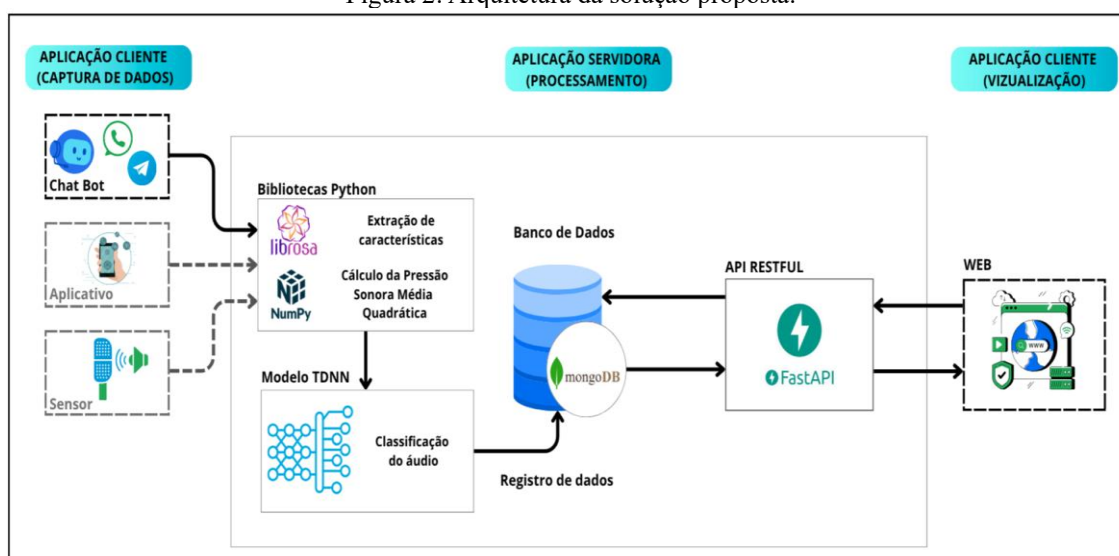


exibição dos mapas de calor, sendo: Aplicação Cliente (Coleta de Dados), Aplicação Servidora (Processamento) e Aplicação Cliente (Visualização).

A Aplicação Cliente voltada à coleta de dados é responsável pela captura de informações brutas de áudio e localização. Conforme ilustrado na Figura 2, essa coleta pode ser realizada por meio de chatbots integrados a aplicativos de mensagens, como o WhatsApp e/ou Telegram, aplicativos dedicados ou sensores específicos. Neste trabalho, foi implementado um chatbot no Telegram. Vale ressaltar, que um chatbot oferece uma forma de coleta de dados não invasiva, uma vez que não exige a instalação de um novo aplicativo. O usuário pode iniciar uma conversa, gravar e enviar um áudio contendo o ruído diretamente pelo chat. Em seguida, é possível compartilhar a localização por meio das funcionalidades nativas da plataforma de mensagens. Essa abordagem pode causar impactos na qualidade de captura dos dados sonoros, uma vez que os aplicativos de mensagens são voltados para gravação de voz humana. Ademais, a qualidade dos microfones dos smartphones também influencia a coleta de dados, visto que esse hardware varia de constituição entre marcas e modelos. Importante destacar que esses impactos não foram objeto de estudos deste trabalho.

A Aplicação Servidora é responsável por receber os dados brutos, extrair informações relevantes, realizar a classificação dos áudios, armazená-los e disponibilizá-los para consulta. Para o processamento dos dados de áudio, utiliza-se a biblioteca Librosa, uma ferramenta em Python especializada em análise de sinais sonoros. Dentre suas funcionalidades, destaca-se a extração de características acústicas, como os níveis mínimos e máximos de pressão sonora (NPS), além da aplicação do filtro de ponderação “A”, que simula a percepção auditiva humana.

Figura 2: Arquitetura da solução proposta.



Fonte: Elaborado pelos autores, 2025.

Complementarmente, a biblioteca NumPy, amplamente utilizada para manipulação de matrizes e vetores, é empregada neste trabalho para calcular a Média Quadrática da Pressão Sonora (Root Mean



Square, RMS). Esse valor representa a energia ou “intensidade” do som, sendo posteriormente convertido para a escala de decibéis (dB), unidade padrão para medição de ruído.

Para a tarefa de classificação dos áudios, é utilizado o modelo de inteligência artificial UrbanSound8K_ECAPA, baseado em Time Delay Neural Network (TDNN). Após a extração das características acústicas, os arquivos de áudio são convertidos do formato .ogg para .wav, compatível com o modelo de classificação adotado. É importante frisar que foi adotada uma taxa de confiança de 60% como critério para validação das classificações.

A escolha desse limiar representa um equilíbrio entre a precisão, a proporção de classificações corretas entre as classificações positivas, e o recall, a capacidade do modelo de identificar todas as amostras relevantes. Um limiar mais alto aumentaria a precisão, todavia poderia descartar classificações corretas com menor confiança. O valor de 60% foi definido para otimizar a captura de eventos sonoros relevantes sem introduzir um número excessivo de falsos positivos. Assim, quando a classificação atinge ou ultrapassa esse limiar, ela é considerada válida; caso contrário, o áudio é classificado como pertencente à categoria “Outros”.

Após a extração das características e a etapa de classificação, é criado um objeto contendo as informações do usuário remetente, as coordenadas geográficas e os dados de áudio. Esse objeto é então registrado em uma instância do MongoDB, um sistema de banco de dados NoSQL que oferece maior flexibilidade no armazenamento e consulta dos dados. Outro componente presente na Aplicação Servidora, desenvolvida por meio da *Framework* FastApi conforme os princípios do Restful, é a Application Programming Interface (API), que atua como ponte de comunicação entre os dados armazenados e aplicações externas. Essa interface contém um endpoint que permite que a Aplicação Cliente (Visualização) acesse os dados processados para exibição em forma de mapas de calor.

6 RESULTADOS E DISCUSSÕES

O primeiro componente da arquitetura, responsável pela captura de dados na aplicação cliente, foi implementado como um chatbot na plataforma Telegram. O sistema realiza o cadastro automático de usuários a partir das informações fornecidas pela própria plataforma, verificando também se o usuário já está presente na base de dados. Além disso, o chatbot gerencia o envio das amostras de áudio juntamente com suas respectivas localizações geográficas.

Adicionalmente, a escolha pelo Telegram como aplicativo de coleta apresentou vantagens em termos de acessibilidade e adesão do usuário, corroborando com a abordagem adotada por Silva e Cunha (2022), que utilizaram chatbot do Telegram para coleta de dados ambientais em comunidades urbanas. Isso reforça o potencial do uso de tecnologias já consolidadas para fins científicos, sem a necessidade de desenvolvimento de aplicações nativas para testar o funcionamento da aplicação, o que reduz custos e acelera o ciclo de prototipagem.



Figura 3: Registro das informações no Banco de Dados.

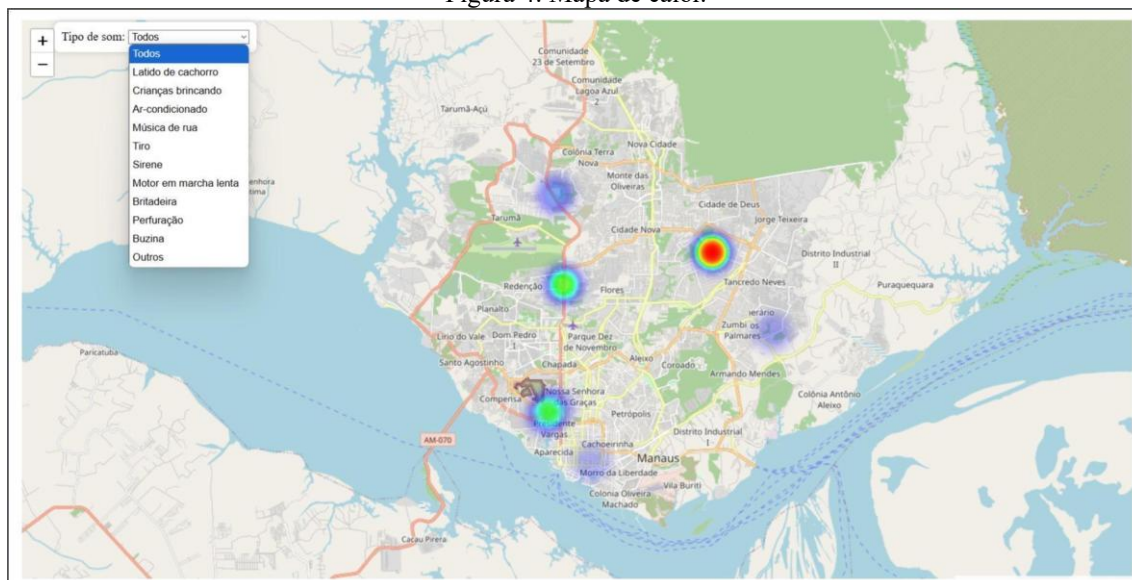
```
_id: ObjectId('6835b7a5642d9163184f8203')
id_user_object: ObjectId('6750565f07ed500dc84f821b')
latitude: -3.0407
longitude: -59.958332
data_criacao: "2025-05-27 09:00:48"
min_db: -80.87924194335938
max_db: -0.8792411088943481
min_dBA: -2.1170837671979967
max_dBA: 77.882916232802
rms_loudness: 0.0429314561188221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

Fonte: Elaborado pelos autores, 2025.

O processamento dos dados mostrou-se eficaz, uma vez que a implementação foi capaz de extrair com precisão (com grau de conformidade entre os dados inseridos) as informações essenciais de cada amostra. Esses dados incluem a classificação do tipo de ruído — realizada por meio do modelo de aprendizado de máquina —, as coordenadas geográficas, a data e a hora do envio. Cada registro consolidado foi armazenado no banco de dados MongoDB Atlas, conforme ilustra a Figura 3.

Com relação à visualização, a Aplicação Cliente consome a API e renderiza um mapa de calor dinâmico, como evidenciado na Figura 4. Esta aplicação possui uma funcionalidade de filtragem por categoria de ruído, permitindo uma análise espacial da poluição sonora de forma específica ou geral do tipo de som.

Figura 4: Mapa de calor.



Fonte: Elaborado pelos autores, 2025.

Diante do exposto, este trabalho mostra a viabilidade do sistema proposto com base em uma arquitetura composta por três aplicações coesas. Essa coesão remete aos princípios da arquitetura baseada em microsserviços, que privilegia serviços independentes e comunicantes por APIs,

promovendo maior modularidade, escalabilidade e resiliência nas aplicações urbanas distribuídas (Ribeiro, Braghetto, 2022). Vale ressaltar que o funcionamento detalhado das aplicações que compõem este sistema pode ser encontrado em (Nery *et al.*, 2025).

7 CONCLUSÃO

Este trabalho consistiu no desenvolvimento de um sistema colaborativo baseado em modelos de aprendizado de máquina para a classificação e o monitoramento georreferenciado de ruídos urbanos, recebidos a partir de dispositivos móveis operados por colaboradores voluntários.

Embora o sistema tenha demonstrado eficácia na correta classificação dos áudios de teste, ainda se faz necessária a realização de experimentos mais amplos, com um conjunto de dados diversificado, a fim de melhorar a acurácia e a generalização do modelo preditivo. Além disso, a etapa de visualização dos dados, embora funcional, ainda passará por reformulações com o intuito de aperfeiçoar a interface gráfica. Tais melhorias visam facilitar o uso por colaboradores e gestores públicos, ampliando o impacto social do sistema. Para futuras melhorias na acurácia da coleta, aponta-se o desenvolvimento de um aplicativo especializado.

Assim, a potencialidade das aplicações desenvolvidas, tanto no lado cliente quanto no lado servidor, indica que a tecnologia possui aplicabilidade em diversos setores da sociedade. Trata-se de uma solução de baixo custo, prática e escalável, cuja implementação pode auxiliar na formulação de políticas públicas, na promoção da saúde coletiva, no suporte à pesquisa científica e na preservação ambiental, especialmente em regiões afetadas por níveis elevados de poluição sonora.

AGRADECIMENTOS

Os autores agradecem ao Instituto Federal de Educação, Ciência e Tecnologia do Amazonas (IFAM) e ao seu Curso Superior em Análise e Desenvolvimento de Sistemas (TADS). Ao Programa de Pós-Graduação em Informática (PPGI) da Universidade Federal do Amazonas e ao Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial (CPGEI-CT) da Universidade Tecnológica Federal do Paraná (UTFPR).



REFERÊNCIAS

- AHMED, A. A.; PRADHAN, B.; CHAKRABORTY, S. et al. Developing vehicular traffic noise prediction model through ensemble machine learning algorithms with GIS. *Arabian Journal of Geosciences*, v. 14, n. 16, p. 1–15, 2021. DOI: 10.1007/s12517-021-08114-y.
- ARINAITWE, Chris; MUGISHA, Samuel; OPIO, Daniel. Review of Techniques used in Speech Signal Processing. *Journal of Signal Processing Systems*, v. 3, n. 1, p. 63–70, 2024. DOI: 10.59568/KJSET-2024-3-1-07.
- BELTRAN, C. A. R. et al. Plataforma de aprendizado de máquina para detecção e monitoramento de alunos com risco de evasão. *Anais do CBIE/SBIE*, v. 30, n. 1, p. 1591, 2019. DOI: 10.5753/cbie.sbie.2019.1591.
- BEZERRA, E. A. A. C. et al. Instrumentos construídos/validados ao processo de enfermagem em saúde mental: revisão integrativa. *Revista Eletrônica Acervo Saúde*, v. 25, p. e19772, 2025. DOI: 10.25248/reas.e19772.2025.
- BEZERRA, E. A. A. C.; FERREIRA, A. A.; PORTELA, L. C. Diagnóstico de enfermagem em saúde mental infantojuvenil. *Miscellaneous*, v. 24, n. 1, p. e13700, 2024. DOI: 10.25248/reas.e13700.2024.
- BOUMCHICH, A. A Smartphone-Based Crowd-Sourced Database for Environmental Noise Assessment: from data quality assessment to the production of relevant noise maps. 2023. 122 f. Tese (Doutorado em Física) – Le Mans Université. Disponível em: <https://theses.hal.science/tel-04444620/>
- CAN, A.; et al. 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*, v. 10, n. 1, p. 20220166, 2023. DOI: 10.1515/noise-2022-0166.
- CERIOTTI, V. C. Classificação de sons ambientais utilizando redes neurais convolucionais para aplicações em hardwares com recursos limitados. 2023. Trabalho de Conclusão de Curso (Graduação) – Universidade Federal do Rio Grande do Sul. Disponível em: <http://hdl.handle.net/10183/267582>.
- CHEN, Kai; KONG, Qiuqiang; IQBAL, Turab; WANG, Yuxuan; WANG, Wenwu; PLUMBY, Mark D. HTS-AT: A Hierarchical Token-Semantic Audio Transformer for Sound Classification and Detection. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022. IEEE, p. 646–650. DOI: 10.1109/ICASSP43922.2022.9746667.
- CHEN, Shuai; PANDA, Asish; HSU, Wei-Ning; MANOCHA, Pranay; SYNNAEVE, Gabriel; MOHAMED Abdelrahman. BEATs: Audio Pre-Training with Acoustic Tokenizers. *arXiv preprint*, arXiv:2212.09058, 2022. Disponível em: <https://arxiv.org/abs/2212.09058>.
- DUBEY, Rakesh; BHARDWAJ, Shruti; SHARMA, Vinamra B.; BHATT, Arpit; BISWAS, Shusham. Smartphone-based traffic noise mapping system. *ISPRS Archives*, XLIII-B4-2022, p. 613–620, 2022. DOI: 10.5194/isprs-archives-XLIII-B4-2022-613-2022.
- GONG, Yuan; CHUNG, Yu-An; GLASS, James. AST: Audio Spectrogram Transformer. *arXiv preprint*, arXiv:2104.01778, v. 3, p. 1–5, 2021. Disponível em: <https://arxiv.org/abs/2104.01778>.
- KONG, Qiuqiang; CAO, Yin; IQBAL, Turab; WANG, Yuxuan; WANG, Wenwu; PLUMBLEY, Mark D. PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition.



IEEE/ACM Transactions on Audio, Speech, and Language Processing, v. 28, p. 2880–2894, 2020. DOI: 10.1109/TASLP.2020.3030497.

KOUTINI, Khaled; SCHLÜTER, Jan; EGHBAL-ZADEH, Hamid; WIDMER, Gerhard. Efficient Training of Audio Transformers with Patchout. In: Interspeech 2022. ISCA, 2022. DOI: 10.21437/Interspeech.2022-227.

LARANJA, Andrea C.; SAITER, Júlia C. Análise de poluição sonora urbana em Vitória–ES a partir do uso de smartphones. ENCAC, v. 17, p. 1–9, 2023. DOI: 10.46421/encac.v17i1.3885.

LIU, Shuai; YUAN, Yi; LIU, Xiulong; LIU, Yan. Audio self-supervised learning: A survey. Patterns, v. 3, n. 12, p. 1–18, 2022. DOI: 10.1016/j.patter.2022.100616.

MACEDO, Alessandra; PERSONA, Leandro; MELONI, Fernando. Recognition of Emotions through Facial Geometry with Normalized Landmarks. In: Proceedings of the 30th Brazilian Symposium on Multimedia and the Web, Juiz de Fora/MG. Porto Alegre: SBC, 2024. p. 257–266. DOI: 10.5753/webmedia.2024.243252.

MEHRISH, Amit; RAJ, Rahul; KUMAR, Sushil; ROY, Partha P. A review of deep learning techniques for speech processing. Information Fusion, v. 99, p. 1–72, 2023. Disponível em: <https://arxiv.org/abs/2305.00359>.

NERY, Sanny C. C.; BATISTA, Fani T. S.; BEZERRA, Sergio A. C. ColabSomGeo: Ferramenta de monitoramento de poluição sonora subsidiada por modelo de aprendizagem. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB, 31º, 2025, Rio de Janeiro. Anais. Local: Sociedade Brasileira de Computação, 2025.

PURWINS, Hendrik et al. Deep Learning for Audio Signal Processing. IEEE Journal of Selected Topics in Signal Processing, v. 13, n. 2, p. 206–219, 2019. DOI: 10.1109/JSTSP.2019.2908700.

RIBEIRO, Murilo B.; BRAGHETTO, Kelly R. A Scalable Data Integration Architecture for Smart Cities: Implementation and Evaluation. Journal of Information and Data Management, v. 13, n. 2, p. 207–223, 2022. DOI: 10.5753/dm.2022.2485.

SANTOS, Alessandro; FREITAS, Leandro de; TEIXEIRA, Igor; VAGNER, Gava; TAIRA, Gustavo; QUILE, Rosa E.; BRAGHETO, Kelly. Desafios e oportunidades da aplicação de Sistemas Ciberfísicos no monitoramento da poluição urbana. 2020. p. 276–289. DOI: 10.5753/courb.2020.12369.

SANTOS, Carlos J.; SOUZA, Luiz E.; PAIVA, Benedito G. M. Análise da influência de diferentes codificadores de voz (CODECs) em redes convergentes. Revista Científica e-Locução, v. 1, n. 17, p. 26–26, 2020. DOI: 10.57209/e-locucao.v1i17.249.

SANTOS, Luís Conde; LEÃO, Cristina; FIGUEIREDO, João; MELO, Afonso; COVAS, Guilherme; CARVALHO, André. PROJETO MIRA – Desenvolvimento de uma Plataforma Inteligente de Monitorização de Ruído Ambiental. 2023. Disponível em: https://www.dbwave.pt/uploads/media_items/mira-article-tecniacustica-20230915-lcs.original.pdf.

SARKER, Iqbal H.; KAYES, A. S. M.; WATTERS, Paul. Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage. Journal of Big Data, v. 6, n. 57, p. 2–28, 2019. DOI: 10.1186/s40537-019-0219-y.



SILVA, Igor S. C.; CUNHA, Felipe. Monitoramento da poluição sonora em Belo Horizonte utilizando redes de sensoriamento participativo. 2022. Disponível em: <https://bib.pucminas.br/pergamumweb/vinculos/00000b/00000bc6.pdf>.

SOFIANOPOULOS, Sofianos et al. Citizens as Environmental Sensors: Noise Mapping and Assessment on Lemnos Island, Greece, Using VGI and Web Technologies. *European Journal of Geography*, v. 15, n. 2, p. 106–119, 2024. Disponível em: <https://www.eurogeojournal.eu/index.php/egj/article/view/476>.

STRAUSS, Edilberto; VILLAS BÔAS JÚNIOR, Manoel; FERREIRA, Wagner L. L. A importância de utilizar métricas adequadas de avaliação de performance em modelos preditivos de machine learning. *Projectus*, v. 7, n. 2, p. 52–62, 2022. DOI: 10.15202/25254146.2022v7n2p52.

TRICCO, A. C. et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, v. 169, n. 7, p. 467–473, 2018. Disponível em: <https://doi.org/10.7326/M18-0850>. Acesso em: 1 nov. 2025.

ZAMAN, Khalid et al. A survey of audio classification using deep learning. *IEEE Access*, v. 11, p. 106620–106649, 2023. DOI: 10.1109/ACCESS.2023.3318015.

ZHAO, Baishan; LIANG, Jiwen. Hierarchical-Concatenate Fusion TDNN for sound event classification. *PLOS ONE*, v. 19, n. 10, p. 1–17, 2024. DOI: 10.1371/journal.pone.0312998.

