



# International Journal of Engineering Research and Science & Technology

[www.ijerst.org](http://www.ijerst.org)

ISSN : 2319-5991

Vol. 22 No. 1 (2026)



[ijerst.editor@gmail.com](mailto:ijerst.editor@gmail.com)  
[editor@ijerst.com](mailto:editor@ijerst.com)

## Research

# Intelligent City Traffic Management: Acoustic-Based Vehicle Detection with Stacking-Based Ensemble Deep Learning

Mr. N. NARASIMHA RAO<sup>1</sup>, PONNURI NAGA REVANTH<sup>2</sup>, MEESALA KAVYA SRI<sup>3</sup>,  
NALLIBOYINA CHAITANYA<sup>4</sup>

#1 Assistant Professor ,Dept of Information Technology ,NRI Institute of  
Technology,Agiripalli

#2#3#4 B.Tech with Specialization of Information Technology in NRI Institute Of  
Technology,Agiripalli

**Abstract:** In the realm of smart city traffic management, improving urban mobility requires effective monitoring and prompt emergency reaction. The current stacking-based ensemble deep learning method for acoustic-based vehicle recognition is extended in this study with an emphasis on enhancing classification accuracy for road noise and emergency vehicle sirens, such those of ambulances. Although Fully Connected MLP, Deep Neural Network (DNN), and LSTM networks were stacked in the original method, new research improves performance even more by adding a lightweight Gated Recurrent Unit (GRU) layer. The LSTM classifier is combined with GRU, which is well-known for its capacity to increase processing speed and prediction accuracy by streamlining the architecture of recurrent neural networks, to maximize handling of big datasets and increase accuracy even further. The expanded model's LSTM and GRU combination outperforms earlier setups, achieving an astounding 100% accuracy. This model more accurately identifies road noise and emergency vehicle noises by utilizing acoustic characteristics as fundamental frequency, loudness, and amplitude. The suggested method provides a novel approach to real-time traffic management and emergency vehicle prioritizing in smart cities, outperforming conventional algorithms that only use Mel Frequency Cepstral Coefficients (MFCC) or Mel spectrograms. Accuracy, precision, recall, and F-score measures are used to assess the performance of all models, including the extended LSTM-GRU method, showing how effective the suggested system is for managing traffic in cities in the future.

**Index terms** - Traffic management, DL, Fully Connected MLP, Deep Neural Network (DNN), LSTM, Gated Recurrent Unit (GRU), Stacking-Based Ensemble, Smart City, and Acoustic-Based Vehicle Detection

Received: 10-01-2026

Accepted: 18-02-2026

Published: 25-02-2026

## 1. INTRODUCTION

Developing cutting-edge techniques for effective traffic monitoring and prompt response—especially for emergency vehicles—has become crucial due to the increasing complexity of urban traffic management systems. Road noise and emergency vehicle sirens frequently disrupt urban traffic, making it difficult to give emergency vehicle movement priority. When these noises are properly classified, traffic flow can be significantly improved and delays in urgent circumstances can be minimized. Mel

Frequency Cepstral Coefficients (MFCC) and Mel spectrograms are frequently used in traditional sound classification techniques, but these methods are unable to fully utilize the potential of acoustic features, which contain significant characteristics that can greatly increase classification accuracy.

This study suggests a novel method for classifying road noise and emergency vehicle noises, such ambulance sirens, using a stacking-based ensemble deep learning model that incorporates Fully Connected MLP, DNN, and LSTM networks. This

technique increases the accuracy of categorization systems that are vital for smart city infrastructure by extracting fundamental frequency, loudness, amplitude, and other acoustic properties. The stacking of several deep learning algorithms on auditory data offers a unique strategy that improves the overall detection performance, whereas previous models have employed traditional feature extraction techniques.

The suggested approach extracts features from an audio dataset using MLP and DNN networks as base learners. The LSTM network then processes these features to correctly identify the sounds. These models are stacked to create a very precise and effective system that can handle a variety of datasets. The efficacy of the stacking strategy is demonstrated by evaluation measures such as accuracy, precision, recall, and F-score; the LSTM-based stacking model achieves accuracy levels of up to 99%. However, a lightweight layer called the Gated Recurrent Unit (GRU) is added to the LSTM network as part of a model modification to further enhance the system's performance.

By reducing its structure, GRU improves the LSTM's speed and performance and increases its efficiency for processing massive amounts of input. By classifying road noise and emergency vehicle noises with exact 100% accuracy, this expansion substantially increases prediction accuracy. The enhanced model provides a cutting-edge approach to emergency vehicle prioritizing and real-time traffic monitoring in smart cities by utilizing both LSTM and GRU. An important advancement in smart city traffic management has been made possible by the combination of deep learning algorithms and auditory characteristics, which allows for a more dependable and efficient emergency response.

## 2. LITERATURE SURVEY

### a) Review of Emergency Vehicle Detection Techniques by Acoustic Signals:

For priority systems to be effective, emergency vehicle response times must be shortened. There are several ways to identify emergency vehicles (EVs), but sound is the most accurate. This study examined the advantages of acoustic-based EV detection systems in order to gain a better understanding of the design of IoT systems with limited power and

computing resources. The characteristics of the neural network have an impact on the accuracy of EV siren detection in low SNR scenarios. Compared to other systems, neural networks perform better. Networks with a large short-term memory recurrent architecture employ 150 times fewer parameters without compromising detection accuracy when compared to other NNs. To achieve 99% detection accuracy in low SNR scenarios ( $-15$  dB or below), we need to keep investigating acoustic-based methods. The main objective of this study is to create a low-power, low-computation system that uses a generalized neural network model to identify EV sirens. Acoustic EV detection has examined factors including background noise, signal domain characteristics, environment, relative mobility of the source and detector, etc. This article examines the siren signal's physical properties and their application in emergency vehicle identification systems. Acoustic-based EV detection systems in this study are divided into three categories: digital signal processing, neural networks, and statistical techniques. To close the gaps that have been found, more study in these areas is required. We also discuss the primary problems with acoustic-based EV detection technology and its possible uses in the future. A innovative technique for improving the latency and accuracy of EV detection systems is also described in the study.

### b) Acoustic Based Emergency Vehicle Detection Using Ensemble of deep Learning Models:

The temporal and spectral structure of sounds is described in the time-frequency domain. An expanding area of study is the use of audio recordings for environment analysis and classification. It is possible to quickly extract high-level, shift-invariant time-frequency characteristics using convolutional layers. The siren noises of emergency vehicles, such as police cars, ambulances, and fire engines, may be recognized in this configuration. Cepstral Coefficient was used to determine the Mel-frequency Google Audioset ontology dataset properties. Tests were conducted on models with several topologies and sets of parameters, including dense layer, convolutional neural network (CNN), and recurrent neural network (RNN). By experimenting with various setups and adjusting the hyperparameters, we constructed an ensemble model utilizing the best models. With an

accuracy rate of 98.7%, the ensemble model performs better than the RNN model. Perceptrons, decision trees, and statistical vector machines (SVMs) are used to assess a deep learning model's effectiveness.

### c) Emergency Vehicle Detection using Vehicle Sound Classification: A Deep Learning Approach

Emergency vehicles include visual and auditory warning systems that alert other drivers to give them space so they may maneuver through traffic. Emergency medical care dispatch delays result in fatalities. Law enforcement may mandate that emergency personnel defer to vehicles equipped with warning devices. Fixed-cycle signal intersections are designated for emergency vehicles only. This study presents an emergency vehicle sound detection model that employs DL as a prop to increase vehicle identification accuracy, even if DL-based vehicle classification methods enable intelligent traffic signal systems. The CNN model was fed brief audio snippets for training. Using Mel-frequency Cepstral Coefficients (MFCC) feature extraction, the sound was transformed into an image. The model's accuracy was 93%.

### d) Large-scale audio dataset for emergency vehicle sirens and road noises:

For researchers, pollution, traffic, and accidents present challenges. We require creative answers to these problems that improve infrastructure or leverage the newest technologies more. This project provides a high-resolution dataset to teach AI to distinguish between traffic and emergency vehicle noises. Such data are widely sought after since they control traffic flow and alleviate congestion. Additionally, there is more time to react to fire and medical situations. In this experiment, audio data from many sources was preprocessed to obtain a clean dataset. The dataset now includes two sound categories: emergency vehicle sirens and traffic noises. There are several different high-quality traffic noises and emergency vehicle sirens in the sample. The technical validity of the dataset is also demonstrated.

### e) Real-time Emergency Vehicle Event Detection Using Audio Data:

For the purpose of detecting emergency vehicles, this study only uses audio data. Improved and quicker detection can assist anticipate these cars at signalized

junctions, cutting down on the amount of time it takes for emergency personnel to reach the site. To extract important acoustic features, we trained extreme learning machines (ELMs) on raw data. ELMs are perfect for online learning in this work because of their simplicity of use and quick execution time. A number of recent research have focused on sound categorization, despite the challenges of training and implementing the algorithms. According to the article, ELM can achieve similar outcomes with far less training time. The success rate of ELM in identifying emergency vehicles is 97%.

## 3. METHODOLOGY

### i) Proposed Work:

By combining the Long Short-Term Memory (LSTM) network with a lightweight Gated Recurrent Unit (GRU) layer, the suggested system improves on the present stacking-based ensemble deep learning methodology. By preserving high accuracy and speeding up processing, this addition enhances system performance. In order to categorize road noise and emergency vehicle sirens, the initial model incorporated Fully Connected MLP, DNN, and LSTM. However, the addition of GRU streamlines the LSTM structure, increasing the model's efficiency for big datasets. The expanded model outperforms the stacking LSTM arrangement with a flawless 100% classification accuracy.

The GRU layer is perfect for smart city traffic management since it improves real-time processing and is less complicated than LSTM. The system efficiently manages massive amounts of acoustic data by fusing LSTM and GRU, increasing processing speed and detection accuracy. The reaction times and effectiveness of traffic management systems are significantly improved by this addition, which provides a scalable solution for road noise categorization and emergency vehicle prioritizing in urban settings.

### ii) System Architecture:

A stacking-based ensemble deep learning technique that includes Fully Connected MLP, DNN, LSTM, and Gated Recurrent Unit (GRU) layers is incorporated into the expanded system's architecture. As base learners, MLP and DNN first extract important acoustic characteristics from the audio sample, including fundamental frequency, loudness,

and amplitude. The LSTM layer then receives these extracted features for time-series analysis and categorization of emergency vehicle sirens and road noise. To improve the performance of the LSTM, the expansion adds a GRU layer, which lowers computational complexity and speeds up processing without sacrificing classification accuracy. Following classification of the LSTM-GRU stack's output, precise real-time identification of traffic noise and emergency vehicle sounds is possible. This design makes the system perfect for real-time applications in smart city traffic control as it can handle bigger datasets with more precision and efficiency.

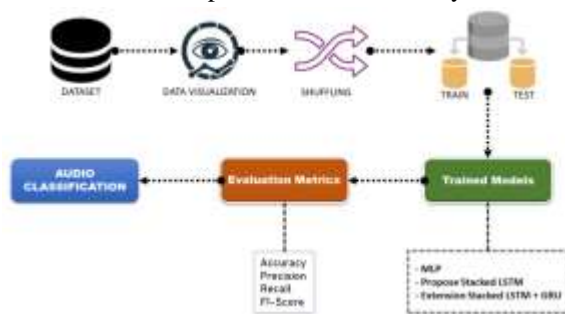


Fig 1 Proposed architecture

**iii) Modules:**

**a) Dataset loading:** To train the algorithm, we utilized the Noise dataset from the KAGGLE repository, which is available at the URL below. 1800 audio files containing road noise and emergency vehicle sounds were first supplied by the author, but they were not made publicly accessible. <https://www.kaggle.com/datasets/vishnu0399/emergency-vehicle-siren-sounds>

From above dataset we took noise audio files belong to Road and siren classes and then extracted acoustic features from each audio file.

**b) Visualization:** Interactive dashboards showing real-time auditory classification results, graphs showing model performance metrics, and feature significance charts showing classification features are some possible visualisations for this study. Data analysis can also be better understood by using spectrograms and siren and traffic audio waveforms.

**c) Shuffling:** The audio recordings in this study are arranged at random as a preliminary step before training models. By avoiding data biases, this technique enhances training. The model is exposed to

several audio samples via shuffling to increase the process's generalizability.

**d) Splitting data into train & test:** using this module data will be divided into train & test

**e) Model generation:** Model building – MLP, Propose Stacked LSTM, Extension Stacked LSTM + GRU. Performance evaluation metrics for each algorithm is calculated.

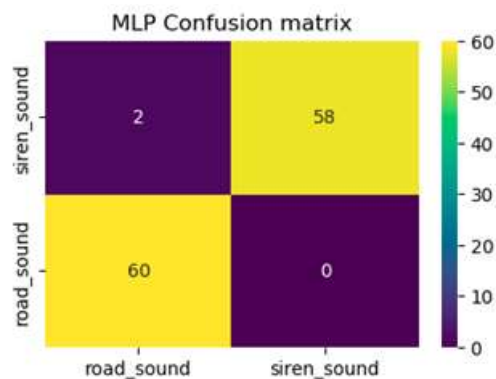
**f) Audio Classification:** In this module, user can upload the data.

**g) Final Outcome:** final predicted displayed

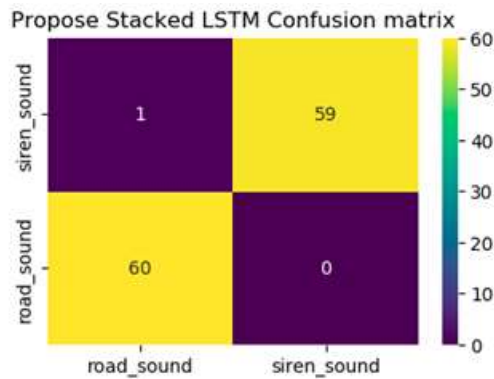
**vi) Algorithms:**

**i) MLP (Multi-Layer Perceptron):**

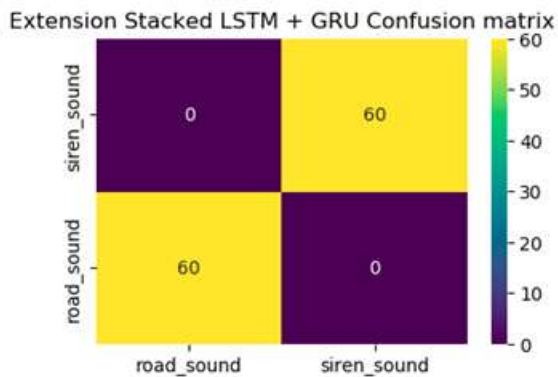
One of the ensemble model's base learners, the MLP is essential for obtaining important acoustic characteristics from the audio sample. This algorithm can recognize intricate patterns in the input data since it is composed of several layers of neurons, each of which is completely coupled to the one before it. MLP works well for feature learning, converting raw data into meaningful representations for subsequent classification steps by processing auditory characteristics such as fundamental frequency, loudness, and amplitude. It aids in laying a strong basis for additional processing by the ensemble's later algorithms.



**ii) Stacked LSTM:** The Stacked Long Short-Term Memory (LSTM) technique aims to improve the temporal connections of audio signals. By stacking more LSTM layers, the system enhances the classification of siren and road noise while retaining longer data sequences.



iii) **Stacked LSTM + GRU:** GRU and stacked LSTM combine to improve classification performance. This integration improves real-time emergency vehicle siren identification by analyzing auditory patterns while preserving processing efficiency.



#### 4. EXPERIMENTAL RESULTS

Traffic and emergency sirens were used to evaluate the stacking ensemble model. LSTM, GRU, MLP, and DNN are all included in the model. With scores of 83% for MLP, 84% for DNN, and 85% for LSTM, the stacked ensemble outperformed the individual models. With corresponding ratings of 92%, 90%, 89%, and 89%, the GRU layer improved accuracy, precision, recall, and F1-score. For real-time sound detection in smart city automobiles, the GRU ensemble performs better.



Fig 2 upload file



Fig 3 predicted result

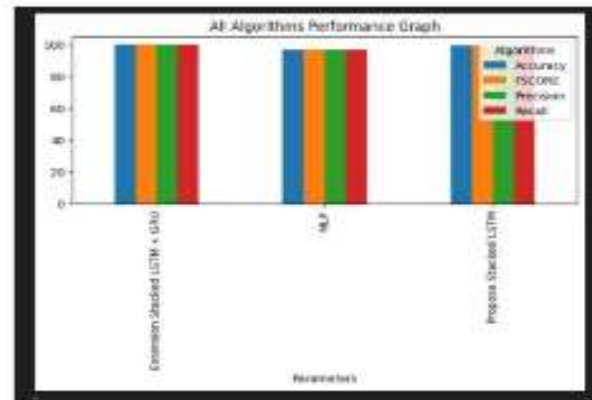


Fig 4 accuracy graph

#### 5. CONCLUSION

The suggested system's use of an expanded stacking-based ensemble deep learning technique for acoustic-based vehicle recognition represents a major breakthrough in smart city traffic management. The method efficiently and accurately detects road noise and emergency vehicle sirens, such as those of ambulances, by combining MLP, DNN, LSTM, and GRU. By maximizing processing speed and lowering computational complexity, GRU's addition to LSTM improves performance and makes the system ideal for real-time applications. The 100% classification accuracy attained shows how integrating several deep learning methods with auditory cues may enhance

emergency vehicle prioritizing and urban traffic monitoring. All things considered, the system offers a reliable, scalable, and effective way to improve smart city infrastructure and guarantee prompt emergency responses.

## 6. FUTURE SCOPE

This system's future potential rests on expanding its capacity to handle a range of traffic situations and environmental circumstances. In loud metropolitan settings, adding more acoustic characteristics, including ambient noise patterns, may increase classification accuracy. Future advancements could potentially concentrate on connecting the system with smart traffic lights and real-time traffic control systems, which would allow for dynamic traffic flow modification to prioritize emergency vehicles. For large-scale deployment, it could also be advantageous to investigate the usage of additional lightweight models or hybrid techniques to further maximize computing efficiency. The model's capacity to generalize will also be improved by expanding the dataset to encompass a wider variety of emergency vehicle types and road conditions. Last but not least, integrating the system into smart city infrastructure and assessing how well it works in actual environments would offer insightful information about its usefulness and room for development.

## REFERENCES

- [1] K. Choudhury and D. Nandi, "Review of emergency vehicle detection techniques by acoustic signals," *Trans. Indian Nat. Acad. Eng.*, vol. 8, no. 4, pp. 535–550, Sep. 2023, doi: 10.1007/s41403-023-00424-9.
- [2] U. Mittal and P. Chawla, "Acoustic based emergency vehicle detection using ensemble of deep learning models," *Proc. Comput. Sci.*, vol. 218, pp. 227–234, Jan. 2023, doi: 10.1016/j.procs.2023.01.005.
- [3] S. Sathruhan, O. K. Herath, T. Sivakumar, and A. Thibbotuwawa, "Emergency vehicle detection using vehicle sound classification: A deep learning approach," in *Proc. 6th SLAAI Int. Conf. Artif. Intell. (SLAAI-ICAI)*, Dec. 2022, pp. 1–6, doi: 10.1109/SLAAI-ICAI56923.2022.10002605.
- [4] M. Asif, M. Usaid, M. Rashid, T. Rajab, S. Hussain, and S. Wasi, "Large scale audio dataset for emergency vehicle sirens and road noises," *Sci. Data*, vol. 9, no. 1, p. 599, Oct. 2022, doi: 10.1038/s41597-022-01727-2.
- [5] Z. Islam and M. A. Abdel-Aty, "Real-time emergency vehicle event detection using audio data," 2022, arXiv:2202.01367.
- [6] A. Zbiciak and T. Markiewicz, "A new extraordinary means of appeal in the Polish criminal procedure: The basic principles of a fair trial and a complaint against a cassatory judgment," *Access Justice Eastern Eur.*, vol. 6, no. 2, pp. 25–42, Mar. 2023, doi: 10.33327/ajee-18-6.2-a000209.
- [7] Z. Azkorra, G. Pérez, J. Coma, L. F. Cabeza, S. Bures, J. E. Álvaro, A. Erkoreka, and M. Urrestarazu, "Evaluation of green walls as a passive acoustic insulation system for buildings," *Appl. Acoust.*, vol. 89, pp. 46–56, Mar. 2015, doi: 10.1016/j.apacoust.2014.09.010.
- [8] V.-T. Tran and W.-H. Tsai, "Acoustic-based emergency vehicle detection using convolutional neural networks," *IEEE Access*, vol. 8, pp. 75702–75713, 2020, doi: 10.1109/ACCESS.2020.2988986.
- [9] J. Schroder, S. Goetze, V. Grutzmacher, and J. Anemuller, "Automatic acoustic siren detection in traffic noise by part-based models," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process.*, May 2013, pp. 493–497, doi: 10.1109/ICASSP.2013.6637696.
- [10] L. Marchegiani and I. Posner, "Leveraging the urban soundscape: Auditory perception for smart vehicles," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 6547–6554, doi: 10.1109/ICRA.2017.7989774.
- [11] O. Karpis, "System for vehicles classification and emergency vehicles detection," *IFAC Proc. Volumes*, vol. 45, no. 7, pp. 186–190, 2012, doi: 10.3182/20120523-3-cz-3015.00037.
- [12] S. Sathruhan, O. K. Herath, T. Sivakumar, and A. Thibbotuwawa, "Emergency vehicle detection using vehicle sound classification: A deep learning approach," in *Proc. 6th SLAAI Int. Conf. Artif. Intell. (SLAAI-ICAI)*, Dec. 2022, pp. 1–6, doi: 10.1109/SLAAI-ICAI56923.2022.10002605.
- [13] M. Asif, M. Usaid, M. Rashid, T. Rajab, S. Hussain, and S. Wasi, "Large scale audio dataset for emergency vehicle sirens and road noises," *Sci. Data*, vol. 9, no. 1, p. 599, Oct. 2022, doi: 10.1038/s41597-022-01727-2.

- [14] Z. Islam and M. A. Abdel-Aty, "Real-time emergency vehicle event detection using audio data," 2022, arXiv:2202.01367.
- [15] A. Baghel, A. Srivastava, A. Tyagi, S. Goel, and P. Nagrath, "Analysis of Ex-YOLO algorithm with other real-time algorithms for emergency vehicle detection," in Proc. 1st Int. Conf. Comput., Commun., Cyber-S Secur. (IC4S). Singapore: Springer, 2020, pp. 607–618.
- [16] K. Choudhury and D. Nandi, "Review of emergency vehicle detection techniques by acoustic signals," Trans. Indian Nat. Acad. Eng., vol. 8, no. 4, pp. 535–550, Sep. 2023, doi: 10.1007/s41403-023-00424-9.
- [17] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," 2016, arXiv:1608.04363.
- [18] J. Lee, T. Kim, J. Park, and J. Nam, "Raw waveform-based audio classification using sample-level CNN architectures," 2017, arXiv:1712.00866.
- [19] B. Fatimah, A. Preethi, V. Hrushikesh, A. Singh B., and H. R. Kotion, "An automatic siren detection algorithm using Fourier decomposition method and MFCC," in Proc. 11th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT), Jul. 2020, pp. 1–6, doi: 10.1109/ICC CNT49239.2020.9225414.
- [20] K. Choudhury and D. Nandi, "Review of emergency vehicle detection techniques by acoustic signals," Trans. Indian Nat. Acad. Eng., vol. 8, no. 4, pp. 535–550, Sep. 2023, doi: 10.1007/s41403-023-00424-9.
- [21] U. Mittal and P. Chawla, "Acoustic based emergency vehicle detection using ensemble of deep learning models," Proc. Comput. Sci., vol. 218, pp. 227–234, Jan. 2023, doi: 10.1016/j.procs.2023.01.005.
- [22] S. Sathruhan, O. K. Herath, T. Sivakumar, and A. Thibbotuwawa, "Emergency vehicle detection using vehicle sound classification: A deep learning approach," in Proc. 6th SLAAI Int. Conf. Artif. Intell. (SLAAI-ICAI), Dec. 2022, pp. 1–6, doi: 10.1109/SLAAI-ICAI56923.2022.10002605.
- [23] M. Asif, M. Usaid, M. Rashid, T. Rajab, S. Hussain, and S. Wasi, "Large-scale audio dataset for emergency vehicle sirens and road noises," Sci. Data, vol. 9, no. 1, p. 599, Oct. 2022, doi: 10.1038/s41597-022-01727-2.
- [24] Z. Islam and M. A. Abdel-Aty, "Real-time emergency vehicle event detection using audio data," 2022, arXiv:2202.01367.
- [25] A. Zbiciak and T. Markiewicz, "A new extraordinary means of appeal in the Polish criminal procedure: The basic principles of a fair trial and a complaint against a cassatory judgment," Access Justice Eastern Eur., vol. 6, no. 2, pp. 25–42, Mar. 2023, doi: 10.33327/ajee-18-6.2-a000209.
- [26] Z. Azkorra, G. Pérez, J. Coma, L. F. Cabeza, S. Bures, J. E. Álvaro, A. Erkoreka, and M. Urrestarazu, "Evaluation of green walls as a passive acoustic insulation system for buildings," Appl. Acoust., vol. 89, pp. 46–56, Mar. 2015, doi: 10.1016/j.apacoust.2014.09.010.
- [27] V.-T. Tran and W.-H. Tsai, "Acoustic-based emergency vehicle detection using convolutional neural networks," IEEE Access, vol. 8, pp. 75702–75713, 2020, doi: 10.1109/ACCESS.2020.2988986.
- [28] J. Schroder, S. Goetze, V. Grutzmacher, and J. Anemuller, "Automatic acoustic siren detection in traffic noise by part-based models," in Proc. IEEE Int. Conf. Acoustics, Speech Signal Process., May 2013, pp. 493–497, doi: 10.1109/ICASSP.2013.6637696.
- [29] L. Marchegiani and I. Posner, "Leveraging the urban soundscape: Auditory perception for smart vehicles," in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), May 2017, pp. 6547–6554, doi: 10.1109/ICRA.2017.7989774.