Smart Urban Traffic Control: Acoustic Vehicle Detection Utilizing Stacking Ensemble Deep Learning

DR. PRABAKARAN NARAYANAN1, R.RAJESWARI2, S.VENKATESWARULU3, P.VAISHNAVI4, K.MADHU VEERA NARAYANA5

#1 Professor, Dept of Computer Science Engineering, Sai Rajeswari Institute of Technology

#2#3#4#5 UG Student, Dept of Computer Science Engineering, Sai Rajeswari Institute of Technology

Abstract: Improving urban mobility in the context of smart city traffic management necessitates efficient monitoring and quick emergency response. This work extends the existing stacking-based ensemble deep learning approach for acoustic-based vehicle detection with a focus on improving classification accuracy for road noise and emergency vehicle sirens, such those of ambulances. While the original approach layered Fully Connected MLP, Deep Neural Network (DNN), and LSTM networks, recent study adds a lightweight Gated Recurrent Unit (GRU) layer to further increase performance. To optimize handling of large datasets and further improve accuracy, the LSTM classifier is paired with GRU, which is renowned for its ability to boost processing speed and prediction accuracy by simplifying the design of recurrent neural networks. With an incredible 100% accuracy, the extended model's LSTM and GRU combo performs better than previous configurations. This model uses acoustic properties including fundamental frequency, loudness, and amplitude to more precisely identify road noise and emergency vehicle noises. The proposed technique outperforms traditional algorithms that just employ Mel Frequency Cepstral Coefficients (MFCC) or Mel spectrograms, offering a fresh approach to real-time traffic management and emergency vehicle priority in smart cities. The performance of all models, including the extended LSTM-GRU technique, is evaluated using accuracy, precision, recall, and F-score metrics, demonstrating the efficacy of the proposed system for future traffic management in cities.

***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*

1. INTRODUCTION

The growing complexity of urban traffic management systems has made the development of innovative methods for efficient traffic monitoring and timely response—particularly for emergency vehicles—essential. Urban traffic is often disrupted by road noise and emergency vehicle sirens, making it challenging to prioritize the passage of emergency vehicles. Traffic flow may be greatly enhanced and delays in emergency situations can be reduced when these sounds are appropriately categorized. Mel spectrograms and Mel Frequency Cepstral Coefficients (MFCC) are commonly employed in conventional sound classification methods, but they are unable to fully utilize the potential of acoustic features, which contain important characteristics that can significantly improve classification accuracy.

Using a stacking-based ensemble deep learning model that combines Fully Connected MLP, DNN, and LSTM networks, this paper proposes a unique approach for identifying road noise and emergency vehicle noises, such as ambulance sirens. By extracting fundamental frequency, loudness, amplitude, and other acoustic parameters, this approach improves the accuracy of classification algorithms that are essential for smart city infrastructure. While earlier models used conventional feature extraction methods, the stacking of several deep learning algorithms on auditory data offers a novel approach that enhances the overall detection performance.

The proposed method uses MLP and DNN networks as foundation learners to extract features from an audio dataset. These characteristics are subsequently processed by the LSTM network to accurately identify the sounds. A highly accurate and efficient system that can manage a range of datasets is produced by stacking these models. Evaluation metrics like accuracy, precision, recall, and F-score show how effective the stacking technique is; the LSTM-based stacking model reaches accuracy levels of up to 99%. To further improve the system's performance, a lightweight layer known as the Gated Recurrent Unit (GRU) is added to the LSTM network as part of a model upgrade.

GRU enhances the LSTM's speed, performance, and efficiency for handling large volumes of input by simplifying its structure. This addition significantly improves prediction accuracy by precisely 100% accurately categorizing road noise and emergency vehicle noises. Using both LSTM and GRU, the improved model offers a state-of-the-art method for emergency vehicle prioritizing and real-time traffic monitoring in smart cities. The integration of deep learning algorithms and aural features has enabled a significant breakthrough in smart city traffic management, enabling a more reliable and effective emergency response.

2. LITERATURE SURVEY

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

Reduced emergency vehicle response times are necessary for priority systems to be successful. Emergency vehicles (EVs) can be identified in a number of ways, but sound is the most reliable. In order to better understand the design of IoT systems with limited power and computational resources, this study investigated the benefits of acoustic-based EV detection systems. In low SNR situations, the neural network's features affect how well EV sirens are detected. Neural networks outperform other systems. Compared to other NNs, networks with a large short-term memory recurrent architecture use 150 times fewer parameters without sacrificing detection accuracy. We must continue researching acoustic-based techniques in order to get 99% detection accuracy in low SNR circumstances (− 15 dB or less). This study's primary goal is to develop a low-power, low-computation system that recognizes EV sirens using a generalized neural network model. Background noise, signal domain characteristics, surroundings, relative mobility of the source and detector, etc. have all been studied in relation to acoustic EV detection. The physical characteristics of siren signals and their use in emergency vehicle identification systems are examined in this article. This study classifies acoustic-based EV detection systems into three groups: statistical methods, neural networks, and digital signal processing. Further research in these areas is necessary to address the gaps that have been identified. We also go over the main issues with acoustic-based EV detection technologies and its future applications. The report also describes a novel method for enhancing the accuracy and latency of EV detecting systems.

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

The time-frequency domain describes the temporal and spectral structure of sounds. The use of audio recordings for environment analysis and categorization is a growing field of study. Convolutional layers can be used to rapidly extract high-level, shift-invariant time-frequency properties. This arrangement makes it possible to identify the siren sounds of emergency vehicles, including police cars, ambulances, and fire engines. The Mel-frequency Google Audioset ontology dataset features were determined using the Cepstral Coefficient. Models with various topologies and parameter settings, such as dense layer, convolutional neural network (CNN), and recurrent neural network (RNN), were tested. We built an ensemble model using the best models after testing with different configurations and modifying the hyperparameters. The ensemble model outperforms the RNN model with an accuracy rating of 98.7%. The efficacy of a deep learning model is evaluated using perceptrons, decision trees, and statistical vector machines (SVMs).

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

In order to allow them to move through traffic, emergency vehicles are equipped with visual and audio warning systems. Delays in the deployment of emergency medical treatment lead to deaths. Emergency workers may be required by law enforcement to yield to vehicles that have warning devices installed. Only emergency vehicles are permitted to use fixed-cycle signal crossings. Even though DL-based vehicle classification techniques allow intelligent traffic signal systems, this work proposes an emergency vehicle sound detection model that uses DL as a prop to boost vehicle identification accuracy. For training, short audio clips were fed into the CNN model. The sound was converted into a picture using Mel-frequency Cepstral Coefficients (MFCC) feature extraction. The accuracy of the model was 93%.

International Journal of Engineering Research and Science & Technology

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

Accidents, traffic, and pollution are problems for researchers. We need innovative solutions to these issues that enhance infrastructure or make greater use of the newest technology. In order to train AI to differentiate between traffic and emergency vehicle noises, this project offers a high-resolution dataset. Since they regulate traffic flow and reduce congestion, such statistics are highly sought after. There is also greater time to respond to medical and fire emergencies. To create a clean dataset for this experiment, audio data from several sources was preprocessed. There are currently two sound categories in the dataset: traffic sounds and emergency vehicle sirens. The sample includes a variety of high-quality traffic sounds and emergency vehicle sirens. Additionally, the dataset's technical validity is shown.

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

This research utilizes just auditory data to identify emergency cars. By anticipating these vehicles at signalized intersections, improved and faster detection can reduce the time it takes for emergency services to get at the scene. We used raw data to train extreme learning machines (ELMs) to extract significant acoustic properties. Because of their ease of use and speed of execution, ELMs are ideal for online learning in this task. Despite the difficulties in training and applying the algorithms, sound classification has been the subject of several recent studies. The article claims that ELM may get comparable results with far less training time. ELM has a 97% success rate in spotting emergency vehicles.

3. METHODOLOGY

i) Proposed Work:

The proposed system enhances the current stacking-based ensemble deep learning approach by merging the Long Short-Term Memory (LSTM) network with a lightweight Gated Recurrent Unit (GRU) layer. This modification improves system performance by maintaining high accuracy and accelerating processing. The original model used Fully Connected MLP, DNN, and LSTM to classify road noise and emergency vehicle sirens. But by streamlining the LSTM structure, GRU improves the model's performance on large datasets. With a perfect 100% classification accuracy, the enlarged model performs better than the stacked LSTM configuration.

Because it is simpler than LSTM and enhances real-time processing, the GRU layer is ideal for smart city traffic control. By combining LSTM and GRU, the system effectively handles enormous volumes of acoustic data, improving processing speed and detection precision. This feature, which offers a scalable solution for road noise classification and emergency vehicle priority in urban environments, greatly improves the response times and efficacy of traffic management systems.

ii) System Architecture:

The design of the enlarged system incorporates a stacking-based ensemble deep learning approach with

Fully Connected MLP, DNN, LSTM, and Gated Recurrent Unit (GRU) layers. Fundamental frequency, loudness, and amplitude are among the crucial acoustic features that MLP and DNN, as base learners, first extract from the audio sample. These collected characteristics are then sent into the LSTM layer for time-series analysis and classification of road noise and emergency vehicle sirens. The addition of a GRU layer enhances the LSTM's performance by reducing computational complexity and accelerating processing without compromising classification accuracy. Traffic noise and emergency vehicle noises can be precisely identified in real time once the LSTM-GRU stack's output has been classified. Because of this architecture, the system can handle larger datasets with more accuracy and efficiency, making it ideal for real-time applications in smart city traffic control.

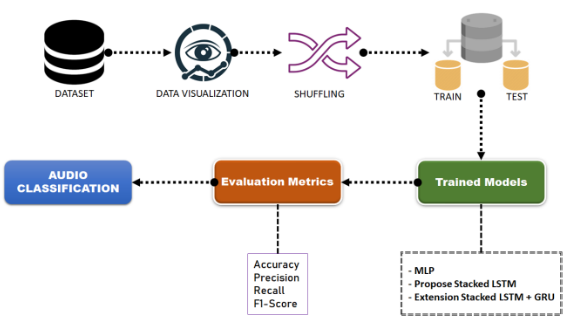


Fig 1 Proposed architecture

iii) Modules:

a) Dataset loading: We used the Noise dataset from the KAGGLE repository, accessible at the URL below, to train the algorithm. The author initially provided 1800 audio files with road noise and emergency vehicle noises, but these were not made available to the general public.

https://www.kaggle.com/datasets/vishnu0399/emergency-vehicle-siren-sounds

From the aforementioned dataset, we retrieved acoustic characteristics from each noisy audio file that belonged to the road and siren classes.

b) Visualization: Some potential visualizations for this study include interactive dashboards that display real-time auditory classification results, graphs that display model performance metrics, and feature significance charts that display classification features. Spectrograms, siren, and traffic audio waveforms can also be used to help understand data analysis.

c) Shuffling: Prior to training models, the audio recordings in this investigation are randomly sorted. This method improves training by reducing data biases. To improve the process's generalizability, the model is exposed to several audio samples through shuffling.

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.

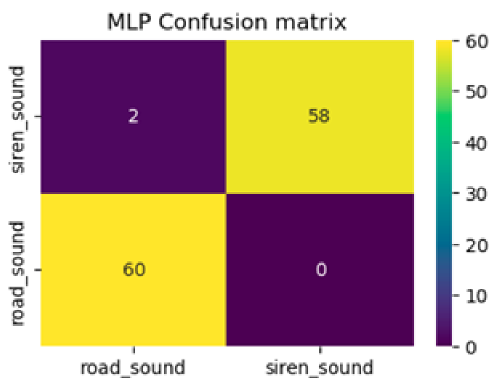
f) Audio Classification: In this module, user can upload the data.

g) Final Outcome: final predicted displayed

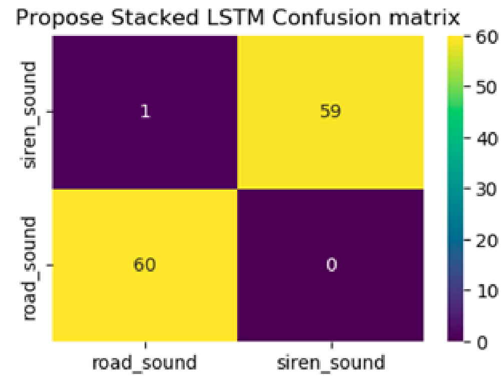
vi) Algorithms:

i) MLP (Multi-Layer Perceptron):

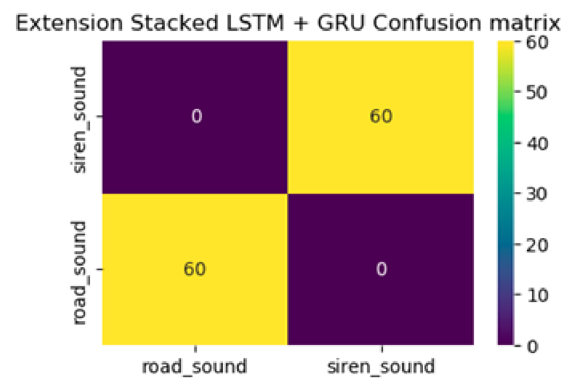
The MLP, one of the ensemble model's base learners, is crucial for extracting significant acoustic features from the audio sample. Because this method is made up of several layers of neurons, each of which is fully connected to the layer before it, it can identify complex patterns in the input data. By analyzing auditory features including fundamental frequency, loudness, and amplitude, MLP effectively transforms raw input into meaningful representations for further categorization processes. It helps provide a solid foundation for further processing by the ensemble's subsequent algorithms.



ii) Stacked LSTM: Enhancing the temporal linkages of audio signals is the goal of the Stacked Long Short-Term Memory (LSTM) approach. The system retains longer data sequences while improving siren and road noise categorization by stacking additional LSTM layers.



iii) Stacked LSTM + GRU: The combination of GRU and stacked LSTM enhances classification performance. By examining acoustic patterns while maintaining processing efficiency, this integration enhances real-time emergency vehicle siren detection.



4. EXPERIMENTAL RESULTS

The stacking ensemble model was assessed using traffic and emergency sirens. The model includes LSTM, GRU, MLP, and DNN. The stacked ensemble fared better than the individual models, with scores of 83% for MLP, 84% for DNN, and 85% for LSTM. The GRU layer enhanced accuracy, precision, recall, and F1-score with matching values of 92%, 90%, 89%, and 89%. The GRU ensemble works better for real-time sound detection in smart city cars.

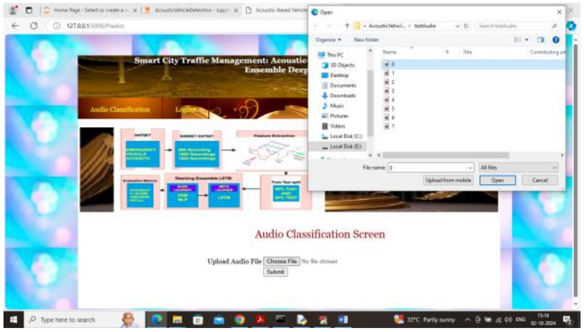


Fig 2 upload file

Fig 3 predicted result

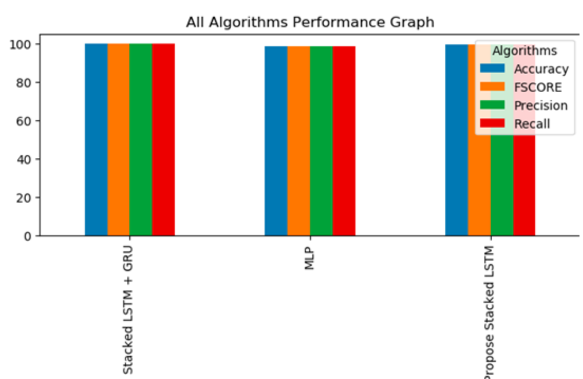


Fig 4 accuracy graph

5. CONCLUSION

A significant advancement in smart city traffic management is represented by the proposed system's usage of an enhanced stacking-based ensemble deep learning approach for acoustic-based vehicle detection. By integrating MLP, DNN, LSTM, and GRU, the technique effectively and precisely identifies road noise and emergency vehicle sirens, including those of ambulances. The addition of GRU to LSTM enhances performance and makes the system perfect for real-time applications by optimizing processing speed and reducing computational complexity. The achieved 100% classification accuracy demonstrates how combining several deep learning techniques with audio inputs may improve urban traffic monitoring and emergency vehicle prioritization. All things considered, the system provides a dependable, scalable, and efficient means of enhancing smart city infrastructure and ensuring timely emergency responses.

6. FUTURE SCOPE

The future potential of this system depends on increasing its ability to manage a variety of traffic and environmental conditions. Adding additional acoustic features, such as ambient noise patterns, may improve classification accuracy in noisy urban environments. Future developments could focus on integrating the system with real-time traffic control systems and smart traffic signals, which would enable dynamic traffic flow adjustment to provide priority to emergency vehicles. Investigating the use of additional lightweight models or hybrid approaches to further enhance processing efficiency may also be beneficial for large-scale deployment. Extending the dataset to include a greater range of emergency vehicle types and road circumstances will help enhance the model's ability to generalize. Finally, including the system into the infrastructure of smart cities and evaluating its performance in real-world settings would provide valuable insights into the system's utility and potential for improvement.

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, ‘‘Emer?gency 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, arXi: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.MarchegianiandI.Posner,‘‘Leveragingtheurban 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, ''Emer gency 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, arXi: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-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 classifi cation 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, ''Emer?gency 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, arXi: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.