URBAN-NET: A LOW LATENCY DEEP LEARNING FRAMEWORK FOR EMERGENCY SIGNAL EXTRACTION IN NOISY ENVIRONMENTS

V. Gayatri[1], K. Dileep[2], Sk. Abdul Muqeem[2], S. Pavan Gowtham Kumar[2], Sk. Muheed[2]

[1]Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (AI & ML)

[1,2]Geethanjali Institute of Science and Technology, Nellore-Bombay Highway, S.P.S.R, Andhra Pradesh 524137, India

To Cite this Article
V. Gayatri, K. Dileep, Sk. Abdul Muqeem, S. Pavan Gowtham Kumar, Sk. Muheed, "URBAN-NET: A LOW LATENCY DEEP LEARNING FRAMEWORK FOR EMERGENCY SIGNAL EXTRACTION IN NOISY ENVIRONMENTS", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 04, April 2026,pp: 989-999, DOI: http://doi.org/10.64771/jsetms.2026.v03.i04.pp989-999

Submitted: 08-03-2026 Accepted: 16-04-2026 Published: 23-04-2026

ABSTRACT

Road accidents account for over 1.19 million deaths globally every year, with delayed emergency response being a major contributing factor. Studies show that automated siren detection can reduce emergency vehicle response time by up to 25 percent in high traffic zones. Smart traffic management systems require real time emergency sound detection to control signals and clear congestion. Surveillance systems, ambulance routing, and disaster response networks rely on accurate siren recognition in noisy outdoor environments. Manual monitoring of audio signals is slow, inconsistent, and highly dependent on human attention. Background noise, overlapping sounds, and fatigue lead to frequent misclassification and delayed emergency response. The proposed system employs a publicly available urban sound dataset consisting of ambulance, fire truck, and traffic audio samples. Audio pre-processing is performed, followed by feature extraction using Mel-Frequency Cepstral Coefficients (MFCC) and Chromagram-based pitch representations (Chroma features) to obtain compact spectral representations. The dataset is divided into training and testing sets for evaluation. Existing classification models include Generalized Learning Vector Quantization (GLVQ), Perceptron, Multi-Layer Perceptron (MLP), and Deep Neural Network (DNN), with a Perceptron-based DNN executed as the primary baseline model. These models are evaluated under noisy urban conditions, demonstrating improved robustness and accuracy for real-time emergency sound detection.

Key words: Smart traffic management, Real time emergency, Congestion control, Rapid urbanization

Disaster response.

1. INTRODUCTION

The response time of vehicles like ambulance, fire truck, and police vehicles are crucial in emergencies. In many cases, it depends on how quickly it can navigate through traffic to reach the destination. Historically, these vehicles have used lights and sirens to alert other road users. Therefore, detection of emergency vehicles using video and audio in real-time can aid in faster navigation especially through intersections. Some intersections are equipped with systems that can aid in emergency vehicle pre- emption with the help of a transmitter in the vehicles that can broadcast information to intersections. A receiver at the intersection is also required for this system. The use of audio data can aid in a much simpler system that could be cheaper and easier to implement on a larger scale. A microphone can be used at the intersection which will be used to distinguish siren from emergency vehicles to provide the right of way. A multi-microphone setup can also be used to understand the direction of the oncoming emergency vehicle.

The figure 1 presents a comparative analysis of A-weighted sound power levels for various emergency sirens across 1/3-octave band centre frequencies ranging from approximately 10 Hz to 10,000 Hz. Each curve corresponds to a different type of siren or emergency service (e.g., police, ambulance, fire brigade, gendarmerie), including both *wail* and *yelp* modes as well as rumblers. Across all curves, the sound power level generally increases from low frequencies, peaks between roughly 700–1500 Hz, and then gradually tapers off beyond 5 kHz. This dominant mid-frequency energy band aligns with human auditory sensitivity, ensuring maximum perceptual clarity in noisy urban environments. Despite slight variations between devices such as the steeper rise in rumblers at lower frequencies or higher peaks for multi-speaker systems the sirens largely follow a similar spectral pattern designed to achieve high detectability.

The overlapping curves in the graph demonstrate that although different countries and manufacturers implement distinct siren designs, they tend to converge around similar sound pressure characteristics to meet international audibility and safety standards. The mid-frequency peak (often approaching or exceeding 110 dB(A)) indicates the region where sirens achieve the greatest acoustic power, optimizing long-distance propagation and penetration through vehicular and environmental noise. Overall, the figure highlights both the consistency and variability in spectral content among global emergency sirens, illustrating how frequency distribution and power levels are engineered to ensure rapid recognition and response in critical situations.

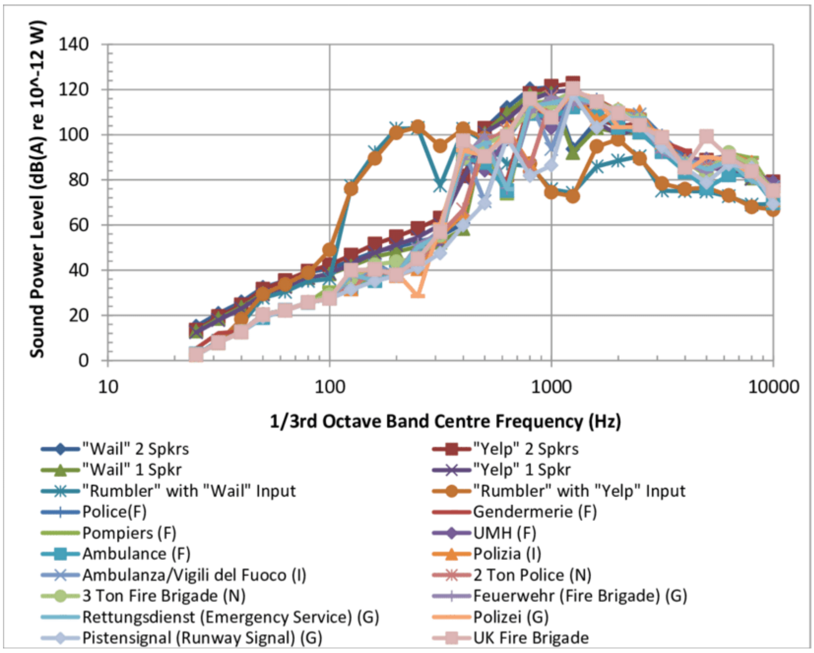

Figure 1. Sound power levels (A-weighted) of all sirens.

In emergency situations, even a delay of a few seconds can be critical. Traditional traffic systems depend heavily on manual control or visual perception by drivers and traffic personnel, which often fails in heavy traffic and noisy urban environments. With the advancement of smart city technologies and Advanced Driver Assistance Systems (ADAS), there is a growing demand for intelligent systems that can sense, analyse, and react automatically to real-world conditions.

Rapid urbanization has significantly transformed modern cities, bringing increased population density, vehicle traffic, construction activity, and industrial operations. While these developments support economic growth, they have also resulted in a dramatic rise in urban noise levels. In dense city environments, continuous background sounds from engines, horns, machinery, and crowds create highly complex acoustic scenes that make it increasingly difficult to identify critical sounds such as emergency vehicle sirens.

2. LITERATURE SURVEY

Smažinka, Dalibor et.al [1] investigated the deployment of intelligent sound event detection (SED) systems capable of recognizing specific sounds, such as gunshots and shouting, within public and commercial spaces. Through controlled simulations in an airport administrative building, the research demonstrates that SED systems significantly outperform traditional notification methods, reducing average response times by over 97% from 175 seconds to just 5 seconds. These findings highlight the potential of SED systems to revolutionize emergency response strategies. The study introduces a novel approach by integrating sound detection with video surveillance into multimodal systems. Rashed, Amr, et.al [2] proposed the gap by contributing in multiple dimensions. Firstly, it emphasizes the significance of sound-based diagnostics for real-time detection of faults through analysing sounds directly generated by vehicles, such as engine or brake noises, and the classification of external emergency sounds, like sirens, relevant to vehicle safety. Secondly, this paper introduces a novel dataset encompassing vehicle fault sounds, emergency sirens, and environmental noises specifically curated to address the absence of such specialized datasetsJayakumar et.al [3] proposed a novel approach to emergency vehicle classification that leverages a comprehensive set of informative audio features to distinguish between ambulance sirens, fire truck sirens, and traffic noise. A unique contribution lies in combining time domain features, including root mean square (RMS) and zero-crossing rate, to capture the temporal characteristics, like signal energy changes, with frequency domain features derived from short-time Fourier transform (STFT).

Badawy et.al [4] introduced an interpretable, sound-based machine learning framework to detect vehicle faults and emergency sound events using acoustic signals as a scalable diagnostic source. Three purpose-built datasets were developed: one for vehicular fault detection, another for emergency and environmental sounds, and a third integrating both to reflect real-world ITS acoustic scenarios. Audio data were pre-processed through normalization, resampling, and segmentation and transformed into numerical vectors using Mel-Frequency Cepstral Coefficients (MFCCs), Mel spectrograms, and Chroma features. To ensure performance and interpretability, feature selection was conducted using SHAP (explain ability), Boruta (relevance), and ANOVA (statistical significance). Joshi et.al [5] investigated in urban city environments; road transportation contributes significantly to the generation of substantial traffic. However, this surge in vehicles leads to complex issues, including hindered emergency vehicle movement due to high density and congestion. Scarcity of human personnel amplifies these challenges. As traffic conditions worsen, the need for automated solutions to manage emergency situations becomes more evident. Intelligent traffic monitoring can identify and prioritize emergency vehicles, potentially saving lives. Zohaib et.al [6] proposed an emergency vehicle detection plays a critical role in ensuring timely responses and reducing accidents in modern urban environments. However, traditional methods that rely solely on visual cues face challenges, particularly in adverse conditions. The objective of this research is to enhance emergency vehicle detection by leveraging the synergies between acoustic and visual information. By incorporating advanced deep learning techniques for both acoustic and visual data, our aim is to significantly improve the accuracy and response times. To achieve this goal, they developed an attention-based temporal spectrum network (ATSN) with an attention mechanism specifically designed for ambulance siren sound detection.

Farooq et.al [7] proposed about designing the correct and efficient real-time system that detects and distinguishes between emergency vehicle sounds so drivers, pedestrians, and also the management systems in their vicinity have prompt recognition and reactions to those sounds. To accomplish this, the proposed solution utilizes acoustic analysis along with sophisticated, cutting-edge algorithms by applying features extraction using Mel-frequency cepstral coefficients (MFCC). Almatawah et.al [8] presented a detailed, data-driven approach for assessing and predicting equivalent continuous noise levels (LAeq) in residential neighbourhoods. The analysis draws on measurements taken at 12 carefully chosen sites covering different road types and urban settings, resulting in 21,720 matched observations. A range of predictors was considered, including road classification, traffic composition, meteorological variables, spatial context, and time of day. Dumitrascu et.al [9] proposed a hardware-constrained Simulink implementation of a yelp siren detector designed for embedded operation. Building on a MATLAB-based proof-of-concept validated in an idealized floating-point setting, the present system reflects practical implementation realities. Key features include the use of a realistically modelled digital-to-analog converter (DAC), filter designs restricted to standard E-series component values, interrupt service routine (ISR)-driven processing, and fixed-point data type handling that mirror microcontroller execution.

Yılmaz et.al [10] proposed, an optimized traffic light system using machine learning that prioritizes the passing of emergency vehicles into city areas. It integrates SVM and Random Forest models by dynamically adjusting traffic light signals based on traffic density to accelerate emergency vehicles. The results reveal that the proposed system would lead to improved emergency response times while enhancing overall transportation efficiency with reduced congestion of traffic. Alam et.al [11] proposed vehicular networks, several automation protocols are invented in artificial intelligence-based systematic processes. But best of our knowledge, none of the methods discussed effective sound detection systems and sound transducers based on real-time scenarios. In this research, an effective sound detection system and sound transducer for an intelligent sound adjustment system in commercial car vehicles using proposed sound prototyping development are developed to create an impact of a sound detection system. This Intelligent sound adjustment system enables six models for vehicle wearable sensor systems with Value Line MSP430 LaunchPad™ Development Kit. Kumar et.al [12] proposed on Criminal incidents such as assault, robbery, and other violent activities pose serious risks to individuals, particularly those who are alone in isolated areas during late hours. Many of these threats are accompanied by distinctive sounds, which can serve as crucial indicators for early detection. Existing security measures often struggle with inefficiencies, such as delays in identifying threats and inaccuracies in classification.

Chun et.al [13] proposed a method for estimating the classes and directions of static audio objects using stereo microphones in a drone environment. Drones are being increasingly used across various fields, with the integration of sensors such as cameras and microphones, broadening their scope of application. Therefore, they suggest a method that attaches stereo microphones to drones for the detection and direction estimation of specific emergency monitoring. Rahman et.al [14] proposed Autonomous unmanned aerial vehicles (UAVs) have several advantages in various fields, including disaster relief, aerial photography and videography, mapping and surveying, farming, as well as defence and public usage. However, there is a growing probability that UAVs could be misused to breach vital locations such as airports and power plants without authorization, endangering public safety. Because of this, it is critical to accurately and swiftly identify different types of UAVs to prevent their misuse and prevent security issues arising from unauthorized access. Mohammed et.al [15] proposed Several assistive technologies (ATs) have been manufactured and tested to alleviate the challenges of deaf or hearing-impaired people (DHI). One such technology is sound detection, which has the potential to enhance the experiences of DHI individuals and provide them with new opportunities. However, there is a lack of sufficient research on using sound detection as an assistive technology, specifically for DHI individuals.

3. PROPOSED SYSTEM

The system focuses on emergency signal extraction in noisy urban environments as shown in Figure 2. using an audio-based classification framework. It processes real-world urban sound data to automatically detect and classify emergency sounds such as ambulance, fire truck, and police sirens, while distinguishing them from traffic noise. The system follows a structured pipeline beginning with audio data acquisition and feature extraction, followed by machine learning and deep learning model training. A Deep Neural Network with a Perceptron-based architecture is proposed as the final model, and its performance is compared with existing classifiers. The trained model is integrated with a Tkinter-based interface to enable real-time prediction and visualization of emergency sound detection results.

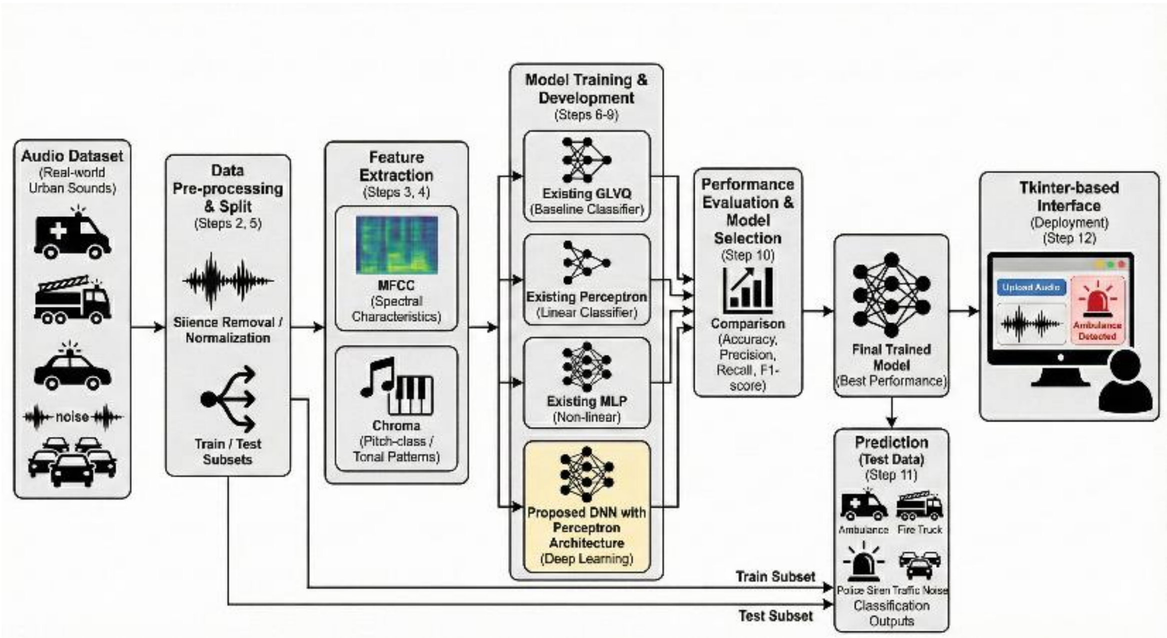

Figure 2. shows the system architecture

The study utilizes a publicly available urban sound dataset comprising labeled audio samples of ambulance, fire truck, police siren, and traffic noise to develop an effective emergency sound classification system. The audio signals are preprocessed through silence removal and amplitude normalization, followed by feature extraction using Mel-Frequency Cepstral Coefficients (MFCC) and Chroma features to capture both spectral and tonal characteristics. The dataset is then split into training and testing sets to ensure unbiased evaluation. Several baseline models, including Generalized Learning Vector Quantization (GLVQ), Perceptron, and Multi-Layer Perceptron (MLP), are implemented for comparative analysis. A proposed Deep Neural Network (DNN) with a Perceptron-based architecture is developed to learn higher-level representations and improve robustness in noisy environments. Model performance is evaluated using accuracy, precision, recall, and F1-score, with predictions generated on unseen test data. Finally, the best-performing model is integrated into a Tkinter-based graphical interface, enabling real-time user interaction for emergency sound detection

DNN with Perceptron

The Deep Neural Network with a Perceptron-based structure as shown in Figure 3. is applied for emergency signal classification to accurately distinguish emergency sirens from traffic noise in complex urban environments. The input to this method consists of MFCC and Chroma-based feature vectors representing ambulance, fire truck, police siren, and traffic sounds. By stacking multiple perceptron layers, the DNN learns hierarchical and discriminative representations that improve robustness against background noise and acoustic variability.

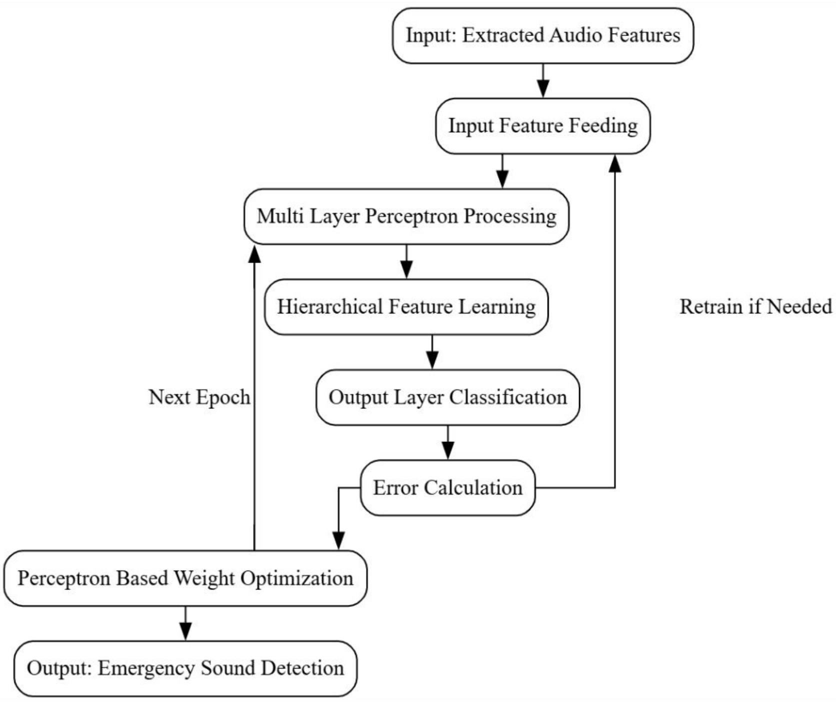

Figure 3. Internal workflow of proposed DNP model

The proposed DNP model begins by feeding extracted audio feature vectors into the input layer, where each feature is mapped to corresponding neurons for further processing. The data is then propagated through multiple hidden layers in a Multi-Layer Perceptron (MLP) structure, where weighted summation and non-linear activation functions enable the learning of complex relationships between acoustic features and sound classes. As the data moves deeper into the network, hierarchical feature learning transforms low-level spectral information into high-level abstract representations that effectively distinguish emergency sounds from traffic noise. The output layer produces class probabilities for each category, indicating the likelihood of the input belonging to a specific class. The predicted outputs are compared with actual labels to compute classification error, which is then backpropagated through the network. Using a perceptron-based weight optimization mechanism, the model iteratively updates its parameters until it achieves stable and accurate emergency sound detection.

4. RESULT ANALYSIS

Figure 4. shows the graphical user interface developed for the proposed emergency signal extraction and classification system, titled "Urban-Net: Emergency Signal Extraction & Classification in Noisy Urban Environments." The interface is divided into two main sections: an Urban Control Panel on the left and a System Intelligence Log on the right. The control panel provides sequential operational buttons including Upload Urban Dataset, MFCC + Chroma Extraction, Train–Test Split, GLVQ Model, Perceptron Model, MLP (DNN), DNN + Perceptron, Live Emergency Prediction, and Exit System, reflecting the complete processing pipeline implemented in the research.

Figure 4. GUI of acoustic event classification in noisy environments

Figure 5. shows the confusion matrix of the proposed DNN with Perceptron model for classifying ambulance, fire truck, and traffic sounds. The results indicate perfect classification of traffic sounds, with all 44 traffic samples correctly identified and no misclassification. The fire truck class also achieves complete accuracy, where all 39 samples are correctly classified as fire truck, demonstrating strong class separability.

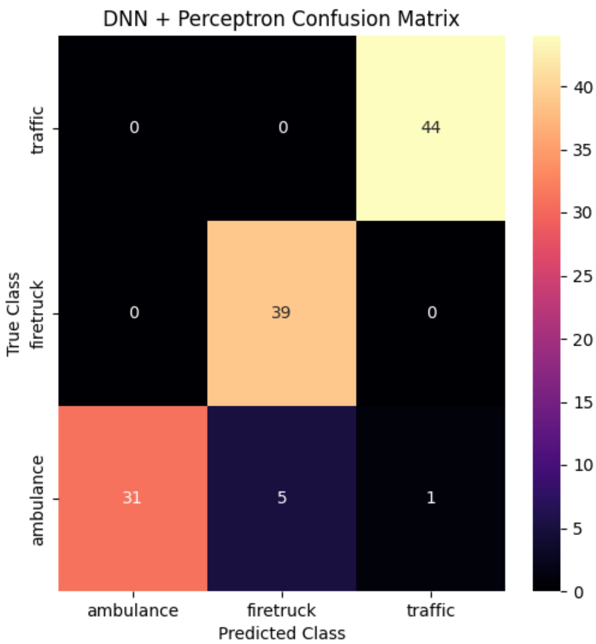

Figure 5. confusion matrix of DNN with Perceptron

For the ambulance class, 31 samples are correctly classified, while 5 samples are misclassified as fire truck and 1 sample is confused with traffic. Compared to GLVQ, Perceptron, and MLP models, this confusion matrix shows significantly reduced inter-class confusion and improved recognition of emergency sirens. Overall, the figure highlights the effectiveness of the proposed hybrid architecture in enhancing classification accuracy and robustness under noisy urban acoustic conditions.

Figure 6. shows the real-time prediction output of the proposed system for an ambulance audio sample, visualized through the integrated graphical user interface. The waveform plot represents the temporal amplitude variation of the input audio signal over a duration of approximately 3 seconds, with amplitude values ranging roughly between −0.08 and 0.06, indicating a high-energy siren pattern typical of emergency vehicles. The system explicitly displays the prediction result as "ambulance" at the top of the waveform, confirming successful classification by the trained DNN with Perceptron model. This figure demonstrates the end-to-end functionality of the system, where an uploaded test audio is processed, classified in real time, and visually presented along with its predicted emergency class, validating the practical applicability of the proposed framework in noisy urban environments.

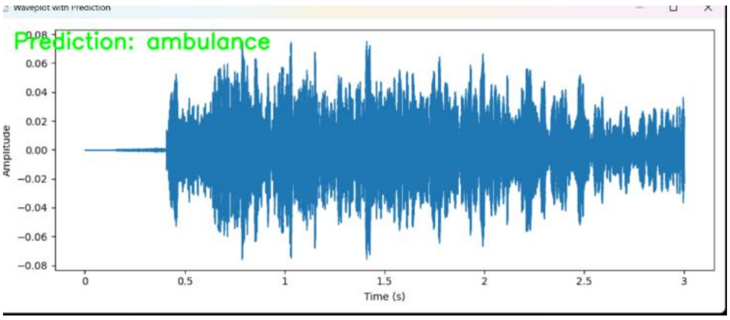

Figure 6. Prediction Results of ambulance class

Figure 7. shows the real-time prediction result of the proposed emergency sound classification system for a fire truck audio sample. The waveform illustrates the amplitude variation of the input signal over a time duration of approximately 3 seconds, with amplitude values fluctuating roughly between −0.6 and 0.6, indicating a high-intensity and continuous siren pattern characteristic of fire truck alarms. The system clearly displays the predicted label as "fire truck," confirming correct classification by the trained DNN with Perceptron model. This figure demonstrates the system's ability to accurately analyze incoming audio, extract relevant features, and provide immediate visual and textual

feedback, thereby validating its effectiveness for real-time emergency sound detection in noisy urban environments.

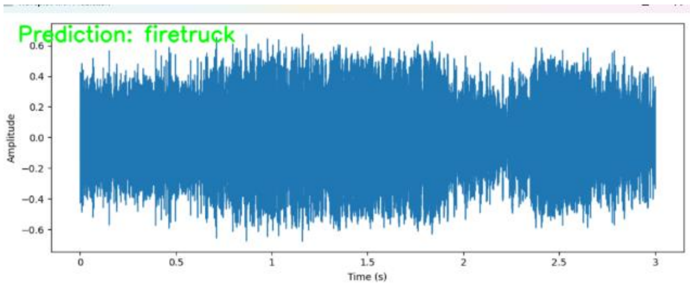

Figure 7. Prediction Results of fire truck class

Figure 8. shows the real-time prediction result of the proposed system for a traffic noise audio sample. The waveform represents the temporal amplitude variation of the input signal over a duration of approximately 3 seconds, with amplitude values ranging roughly between −0.4 and 0.3, reflecting the irregular and non-tonal nature of urban traffic noise. The system clearly displays the predicted class as "traffic," confirming accurate identification by the trained DNN with Perceptron model. This figure demonstrates the system's capability to distinguish non-emergency traffic sounds from emergency sirens, validating its robustness and reliability for real-time operation in noisy urban environments.

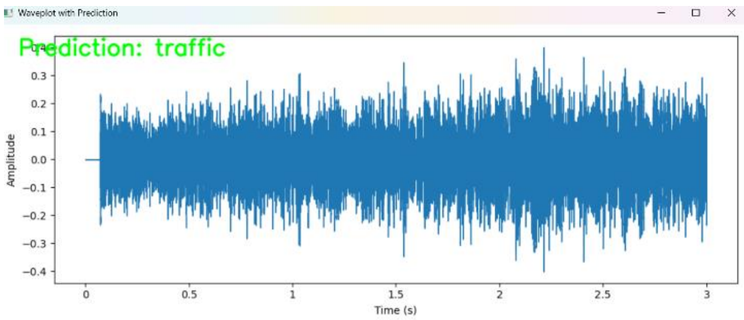

Figure 8. Prediction Results of traffic class

5. CONCLUSION

The experimental evaluation of the proposed emergency signal extraction and classification system demonstrates clear and consistent performance improvements over existing methods. Baseline models such as GLVQ and the traditional Perceptron achieved accuracies of 87.50% and 89.17% respectively, showing limitations in discriminating acoustically similar emergency sirens. The MLP classifier improved overall performance to 91.67% accuracy with a high ROC–AUC value of 98.82%, confirming strong non-linear learning capability. The proposed DNN with Perceptron model delivered the best results, achieving an accuracy of 95.00%, precision of 95.47%, recall of 94.59%, and F1-score of 94.68%. Class-wise analysis further confirmed perfect traffic detection (100% recall) and complete fire truck recognition (100% recall), while significantly reducing ambulance misclassification. These results validate the effectiveness of combining deep feature extraction with a Perceptron-based decision layer for robust emergency sound classification in noisy urban environments. In addition to quantitative performance gains, the system successfully demonstrated real-time operability through a Tkinter-based graphical interface. The GUI-enabled workflow supported dataset upload, feature extraction, model training, comparative evaluation, and live emergency prediction within a unified platform. Real-time prediction results for ambulance, fire truck, and traffic audio samples confirmed correct classification and intuitive waveform visualization, highlighting the system's practical applicability. Overall, the proposed framework achieves a strong balance between accuracy, computational efficiency, and usability, making it suitable for integration into intelligent transportation systems and smart city emergency response infrastructures.

REFERENCES

[1] Smažinka, Dalibor, Radomír Ščurek, and Martin Hrinko. "Accelerating Emergency Response in Airport Environments: An Experimental Study on Intelligent Sound Detection Systems." International Journal of Safety & Security Engineering 15, no. 4 (2025).

[2] Rashed, Amr, Yousry Abdulazeem, Tamer Ahmed Farrag, Amna Bamaqa, Malik Almaliki, Mahmoud Badawy, and Mostafa A. Elhosseini. "Toward Inclusive Smart Cities: Sound-Based Vehicle Diagnostics, Emergency Signal Recognition, and Beyond." Machines 13, no. 4 (2025): 258

[3] Jayakumar, Dontabhaktuni, Modugu Krishnaiah, Sreedhar Kollem, Samineni Peddakrishna, Nadikatla Chandrasekhar, and Maturi Thirupathi. "Emergency Vehicle Classification Using Combined Temporal and Spectral Audio Features with Machine Learning Algorithms." Electronics 13, no. 19 (2024): 3873.

[4] Badawy, Mahmoud, Amr Rashed, Amna Bamaqa, Hanaa A. Sayed, Rasha Elagamy, Malik Almaliki, Tamer Ahmed Farrag, and Mostafa A. Elhosseini. "From Sensors to Insights: Interpretable Audio-Based Machine Learning for Real-Time Vehicle Fault and Emergency Sound Classification." Machines 13, no. 10 (2025): 888.

[5] Joshi, Aruna Kumar, and Shrinivasrao B. Kulkarni. "Multi-modal information fusion for localization of emergency vehicles." International Journal of Image and Graphics 25, no. 05 (2025): 255005

[6] Zohaib, Muhammad, Muhammad Asim, and Mohammed ELAffendi. "Enhancing emergency vehicle detection: A deep learning approach with multimodal fusion." Mathematics 12, no. 10 (2024): 1514.

[7] Farooq, Hira, Muhammad Shadab Alam Hashmi, Talha Farooq Khan, Qamar Hafeez, and Muhammad Mohsin. "Intelligent emergency vehicle sound classification for public safety." Kashf Journal of Multidisciplinary Research 1, no. 12 (2024): 141-152.

[8] Almatawah, Jamal, Mubarak Alrumaidhi, Hamad Matar, Abdulsalam Altemeemi, and Jamal Alhubail. "An Interpretable Machine Learning Framework for Urban Traffic Noise Prediction in Kuwait: A Data-Driven Approach to Environmental Management." Sustainability 17, no. 19 (2025): 8881.

[9] Dumitrascu, Elena Valentina, Răzvan Rughiniș, and Robert Alexandru Dobre. "Design and Evaluation of a Hardware-Constrained, Low-Complexity Yelp Siren Detector for Embedded Platforms." Electronics 14, no. 17 (2025): 3535.

[10] Yılmaz, Yıldıran. "Machine Learning-Enhanced Traffic Light Optimization System Prioritizing Emergency Vehicle Passage Using SVM and Random Forest Models." Gazi University Journal of Science Part A: Engineering and Innovation 12, no. 1 (2025): 175-196.

[11] Alam, Shadab, Omer K. Jasim Mohammad, Badria Sulaiman Alfurhood, Kuldeep K. Saxena, Anand M, R. Mahaveerakannan, and V. Savitha. "Effective sound detection system in commercial car vehicles using Msp430 launchpad development." Multimedia Tools and Applications 84, no. 29 (2025): 35443-35468.

[12] Kumar, S. Tharun, V. Soumya Sree, V. Praveen, V. Veera Chandrika, T. Naveendra, and P. Chandra Sekhar. "ECHOSHIELD:'Turning Sound into Safety'An AI-Powered Sound Detection and Safety Alert System." International Journal of Human Computations & Intelligence 4, no. 3 (2025): 465-475.

[13] Chun, Chanjun, Hyung Jin Park, and Myoung Bae Seo. "Static Sound Event Localization and Detection Using Bipartite Matching Loss for Emergency Monitoring." Applied Sciences 14, no. 4 (2024): 1539.

[14] Rahman, Md Habibur, Mohammad Abrar Shakil Sejan, Md Abdul Aziz, Rana Tabassum, Jung-In Baik, and Hyoung-Kyu Song. "A comprehensive survey of unmanned aerial vehicles detection and classification using machine learning approach: Challenges, solutions, and future directions." Remote Sensing 16, no. 5 (2024): 879.

[15] Mohammed, Hassan BM, and Nadire Cavus. "Utilization of Detection of Non-Speech Sound for Sustainable Quality of Life for Deaf and Hearing-Impaired People: A Systematic Literature Review." Sustainability 16, no. 20 (2024): 8976.