2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM) | 979-8-3503-6509-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICONSTEM60960.2024.10568592

Automatic Recognition of Ambulance Siren by Traffic Signal

Yuvarani P, Suvetha C, Shree Subha M, Subashri V, Leena G

Department of ECE, M Kumarasamy College of Engineering, Karur, India.

E-mail : yuvaranip.eie@mkce.ac.in, suvethasekar2002@gmail.com, shreesubhamammu@gmail.com, subashrivadivelu@gmail.com, vijileena5@gmail.com

Abstract- The most serious issue that densely populated cities confront is traffic congestion. This effort primarily intends to provide a solution to the difficulty encountered by ambulances while approaching a traffic signal amid heavy traffic. This is accomplished by employing an innovative approach for enhancing emergency response through the integration of automatic ambulance siren recognition with traffic signal control systems. Using advanced audio processing and deep learning techniques, the system identifies the unique acoustic profile of ambulance sirens in real-time. This project involves creating a real-time system using Tensor Flow, a machine learning framework, to distinguish ambulance sounds from other noises. Through machine learning techniques, the system will learn to classify audio samples, identifying the unique siren patterns of ambulances even in the presence of ambient noise. When a siren is detected, the traffic signal dynamically adjusts to prioritize the ambulance's passage, minimizing response times and improving overall emergency vehicle maneuverability in urban environments. The objective is to develop a dependable system that accurately recognizes ambulance sounds, potentially enhancing emergency response and bolstering public safety measures. The proposed solution offers a practical and efficient way to optimize traffic flow during emergencies, contributing to enhanced public safety and potentially saving crucial time during emergency situations.

Keywords —Traffic congestion, Ambulance, Deep Learning, Traffic signal

I. INTRODUCTION

The number of cars utilized by people is continually rising due to fast population expansion, which causes high density traffic and increases vehicle waiting time. Ambulances, fire and disaster prevention vehicles, and VIP automobiles must arrive at their location as soon as possible. This system focuses on offering a smart means to regulate traffic as ambulances approach the signal. This work is done by classifying and extracting the audio features and automatically recognizing the ambulance siren.

Fig.1 . Emergency Vehicle Stuck in Traffic

The audio is recognized, the result is inputted to an Arduino UNO platform. If the result predicted is the recognition of ambulance siren, then the relevant lane becomes green, allowing the ambulance to travel entirely through the lane. Developing nation's population, such as India's, are fast growing. The number of autos will increase in lockstep with population growth. As a result, there are several concerns to be concerned about, such as considerable traffic congestion and traffic law violations, among others. Generally, traffic congestion is associated with a plethora of additional issues, such as emergency vehicles becoming stuck. The time it takes for emergency vehicles to arrive has significantly increased. As a result, given the current traffic condition, supporting an emergency vehicle in getting out of a traffic jam is critical.

II.OBJECTIVE :

- To investigate and catalogue the various kinds of signal pre-processing methods that may be applied to create a reliable dataset that the deep learning algorithm can utilize.
- To gather and get data by using various pre-processing techniques in order to produce distinct training and test datasets for the deep learning algorithm.
- Using the pre-processed datasets as deep learning algorithm test cases and keeping track of the algorithm's performance.

- To evaluate and compare the correctness of the findings acquired and record the effectiveness of various preprocessing methods employed to produce the desired outcome.
- To create a report that records and summarizes the project details, implementation and results obtained.

III. LITERATURE SURVEY

Using Vibrionic Coherence Maps Generated by Stimulated Ultrafast X-Ray Raman to Visualize Conical Intersection Passages, Daniel Keefer et al. [1] state. This publication offers a thorough description of the subject of ultrafast X-rays. Curved junctions affect almost all photophysical and photocatalytic processes' speeds and outcomes. On the nuclear landscape of molecules, there are areas of moral decay between electronic states where electrons and nuclei grow on similar timescales and become intimately associated, allowing due to electron relaxation pathways upon visual stimulation. The great speed and complexity of conical junctions make experimental monitoring of them a challenging issue. To offer a simulation study based on a quantum description of the nucleus for the ultrafast photo-relaxation of purine. By capturing the transient wave packet coherence during this transit with an X-ray free electron laser pulse, they are able to show an extra window into conical junctions. It presents two noteworthy findings. Initially, we discover that it is possible to quantify the vibronic coherence at the conical junction for a few hundred femtoseconds. Second, the Wigner spectrograms of the signal immediately yield the time-dependent energy-splitting topography of the contributing electronic and vibrational states. The proposed experiment directly maps the track of a nuclear wave packet at the conical junction. This work provides a detailed explanation of the vehicle detection process used in the vision-based vehicle detection and counting system proposed by Liang et al. [2] in highway settings. Detecting vehicles using intelligence becomes much important in managing highways. But because automobiles come in a range of sizes, it could be challenging to recognize them, which affects how precisely vehicle counts are performed. The recommended dataset provides the whole set of data needed for deep learning-based vehicle recognition. In the proposed vehicle verification and counting system, the recently published segmentation approach—which is crucial for improving vehicle recognition—first removes the highway road surface from the image and splits it into a distant region and a proximate area. The kind and position of the vehicle are then ascertained via the Yolov3 network by utilizing the two zones that were previously specified. The vehicle trajectories, which are ultimately produced by the ORB algorithm, can be utilized to ascertain the number of cars and the direction in which each vehicle is traveling. The results of the experiment lend credence to the idea that implementing the recommended sedimentation approach can improve detection accuracy, especially when searching for little vehicle items. The major objective of this study is to recognize auto accidents in video, as suggested by V.Machacha etal. [3] in their proposal for intelligence fast car crash detection in video. They provide a three- stage structure: The first stage uses convolutional neural networks to detect cars; in this case, the You Only Look Once (YOLO) network was used; the second stage uses a tracker to focus on each car; and the third and final stage uses a Support Vector Machine (SVM) in conjunction with the Violent Flow (ViF) descriptor for each car to detect car crashes. With only 0.5 seconds of latency, their concept operates almost in real time, and they have achieved an accuracy of 89 percent in identifying car crashes Florian Damerow et al. have suggested the Intersection Warning System for Occlusion Risks Using Relational Local Dynamic Maps [4]. The main objective of this study is to address the challenge of risk assessment in traffic settings where there is limited sensory coverage and poor observability. Here, they concentrate on intersection scenarios when visual perception is difficult. They find the field of sight on a local dynamic map that contains geometrical data and road infrastructure by using ray casting. Based on the area with *reduced* visibility, they first generate scene components that potentially offer a risk but are not yet visually detectable. Next, we forecast a worst-case trajectory by estimating collision risk using the survival analysis. The danger indications that are produced can then be utilized to evaluate the driver's behaviour at that precise moment, notify the driver in an emergency, counsel the driver on appropriate behaviour, or plan safe routes. We confirm our methods by utilizing the resulting intersection warning system in practical settings.

IV. METHODOLOGY

I. EXISTING METHODOLOGY

The methodology that exists says that ambulance is detected by using image processing with the help of a deep learning technique CNN and YOLO.

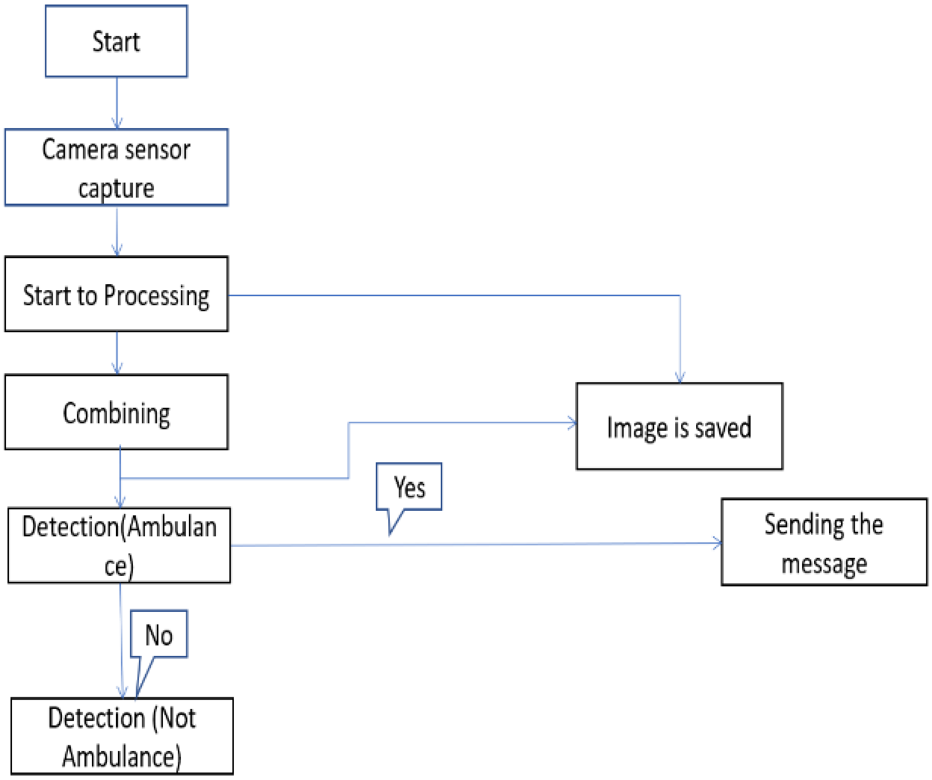

Fig. 2 . Flow Chart of Existing Methodology

This existing method creates a collection of photos of emergency vehicles such as fire truck, police cars and ambulances and the dataset ought to be varied and inclusive of various situations and illumination levels. Next, it preprocess the data by dividing the dataset into training and testing sets, standardizing the pixel values, and resizing the photos to a consistent size. Finally, it creates a model that can identify emergency vehicles in traffic signals and create green corridors for them using the Flask framework, CNN Xception architecture with RMSprop optimizer, and transfer learning. The below figure shows how the existing methodology works. This methodology may capture the images of emergency vehicles in both normal and emergency situations.

V. PROPOSED METHODOLOGY

A. CNN WORKFLOW MODEL

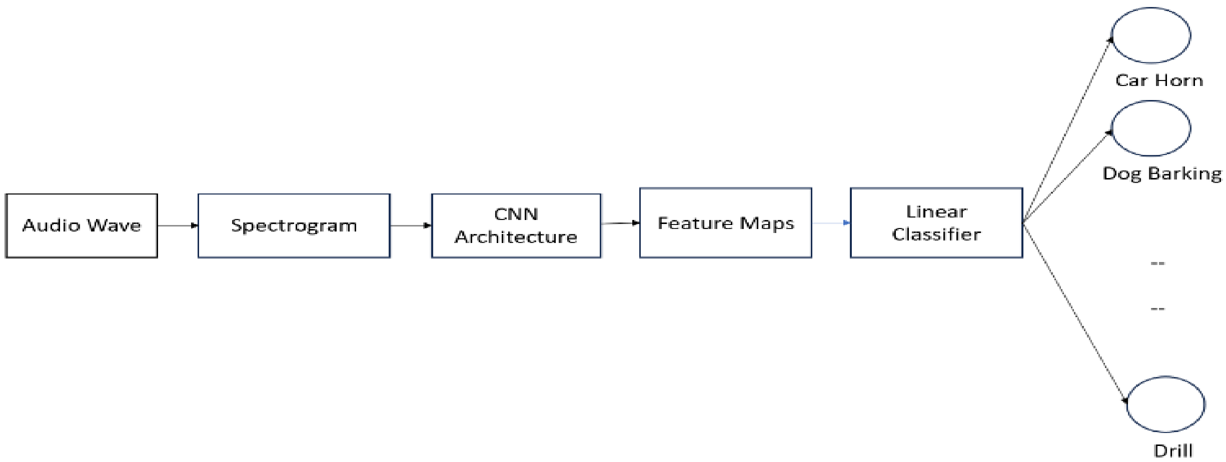

Fig.3 . CNN audio classification work flow

Spectrogram : Converts an audio waveform into a spectrogram involves dividing the waveform into overlapping sections, smoothing them, using Fourier Transform to analyze the frequency components, and creates a picture that illustrates the frequencies and their intensity over a time period.

CNN Architecture : Using a spectrogram, the CNN audio classification framework examines audio data. The input layer, convolutional layer, Relu function, and pooling layers make up its four primary layers. The convolutional layer finds local patterns, while the input layer receives spectrogram data as input. The activation function is the Relu function. By making the feature maps smaller, the pooling layers assist CNN in concentrating on key features. In order to classify the audio, the CNN next compresses the remaining data into a one-dimensional format and runs it through a few more final layers

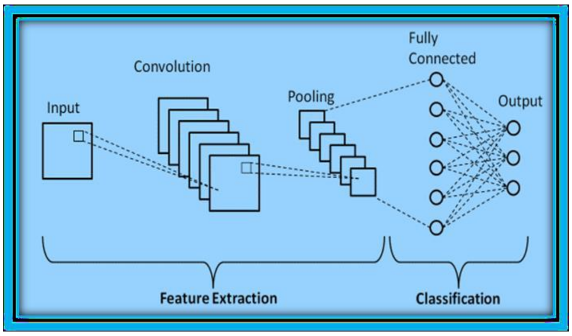

Fig.4 . CNN Architecture outline

Linear Classifier : After reducing the dimensionality of the features, an audio classification algorithm is employed to categorize the audio into predetermined classes. Linear classifiers like Support Vector Machines (SVMs) and logistic regression are frequently utilized for this audio classification task due to their simplicity and efficiency.

Final Output : The audio classification model gives a prediction of the type of sound in an input audio sample, such as a car horn, dog barking, engine noise, or drill sound, based on the relationships it has learned between input features and class labels.

B. TRAINING AND TESTING

Training Flow - The first step in the suggested technique is to input the dataset of the siren or other vehicle sounds. After that, the audio files given as the dataset is collected and pre-processed (filtration of noise) by data acquisition. The characteristics or features of the siren sound or other traffic or vehicle sounds are now extracted using the filtered audio. The dataset that is provided as the input can be used to extract these features. Subsequently, the traits that were extracted are categorized into several types of sounds, including typical car noises, horn sounds, ambulance sirens etc., using Tensor flow machine learning framework. The Keras model, a machine learning library model, is used for this classification. By transforming audio files into spectrograms, Keras model facilitates the loading and preprocessing of audio data. Even with limited datasets, Keras can substantially enhance the performance of audio categorization models.

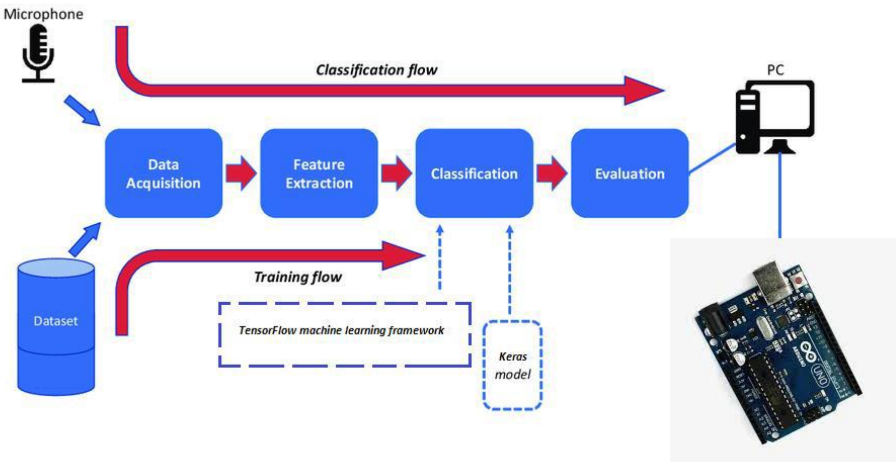

Fig.5 . Block Diagram

Testing flow - The first step in the suggested technique is to use a microphone to record the sound of the siren or other vehicle horn sound. The same process is done as in the training flow until classification. Now then, if the input audio recorded by the microphone is evaluated to check the matching features of the trained dataset. The evaluation that fits the dataset to the classified sound by training flow is now completed by utilizing the evaluation. And so, when the sound of the ambulance siren is identified, the outcome is displayed on the PC as the number “ambulance detected” - 1 or “ambulance not detected” - 0 .

C. WORKING

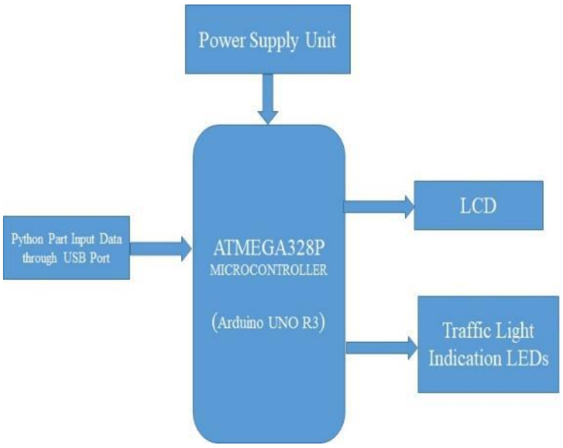

Fig.6 . Embedded Block Diagram

The suggested system's outcome is entirely achieved with Python, and the predicted output from the evaluation is provided to the Arduino UNO (microcontroller) as input data via a USB port. An external power supply may also be used, depending on the needs. The code is embedded into the Arduino to dynamically adjust the traffic LED in case of ambulance siren recognition. After reading the input, if the value at the serial monitoring is “1” which is shown on the LCD, the traffic light LED is changes to GREEN if the result indicates that the ambulance siren has been detected; if not, the traffic signal operates at RED.

D. OUTCOME

If traffic signals can detect ambulance sirens, they could prioritize the ambulance by adjusting the signal to allow the emergency vehicle to pass through quickly. This can potentially save crucial time during emergencies. With quicker passage through intersections, ambulances can reach their destinations faster, potentially saving lives by providing timely medical assistance. Such a system might also contribute to better traffic flow by minimizing disruptions caused by emergency vehicles, reducing congestion, and preventing accidents due to sudden maneuver by drivers trying to clear the way for ambulances.

VI. Results And Discussion

This suggested solution uses the Convolutional Neural Network (CNN), a deep learning algorithm to achieve excellent accuracy in recognizing the presence and absence of ambulance sirens for samples of real-world traffic sounds.

TABLE I . RESULTS OF AUDIO RECOGNITION BY CNN ALGORITHM

Input	Predicted output	Result
Siren audio file	The predicted class is ambulance	1
Other audio file (i.e., traffic noise, dog bark, street music, car horn etc,.)	The predicted class is traffic sound	0

A. MFCC RSEULT OF SIREN

In this proposed innovative approach, one type of supervised learning task is speech recognition which is done using CNN. A popular feature extraction method in speech and audio processing is called MFCC. The spectrum properties of sound are represented by MFCCs. This study first used the feature.mfcc function of the Librosa library to extract Mel Frequency Cepstral Coefficients(MFCC) from the sound dataset since the CNN model is the most appropriate for audio recognition and classification of image. It provides an overview of the correlation between the measured frequency values and the perceived frequency range of the recorded audio. These aural depictions enable us to discern characteristics necessary for categorization.

```
audio_path = audio_dataset_path + "fold1/106905-8-0-0.wav"
(xf, sr) = librosa.load(audio_path)
mfccs = librosa.feature.mfcc(y=xf, sr=sr, n_mfcc=40)
librosa.display.specshow(mfccs, x_axis="time")
plt.colorbar()
plt.tight_layout()
plt.title("MFCC Of Siren")
plt.show
```

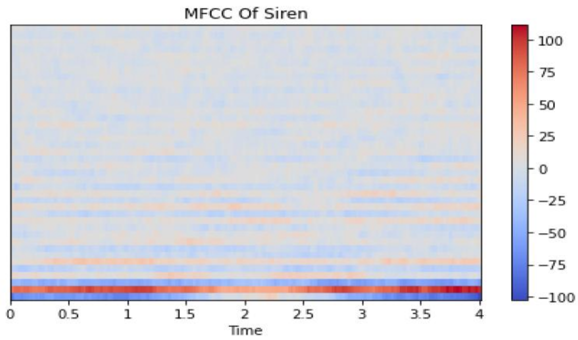

Fig.7 . MFCC of siren

The above figure shows the MFCC result of siren. MFCCs are a collection of coefficients that represent the form of a sound signal's power spectrum.

B. ACCURACY & LOSS GRAPH OF CNN MODEL

Accuracy and Loss is the trained value whereas the validation accuracy and validation loss is the testing value. Convolutional Neural Network attains almost 93% of accuracy with minimum loss. CNN model works more efficient than other algorithms with better accuracy for audio recognition. The model accuracy and loss graph of

CNN are shown in Fig 8 & 9.

Accuracy – Shows the effectiveness of how well the classification model performs or it is able to predict the true output.
Validation accuracy – Shows the accuracy for the new dataset.

The below figure shows that the accuracy is higher than validation accuracy for each epochs.

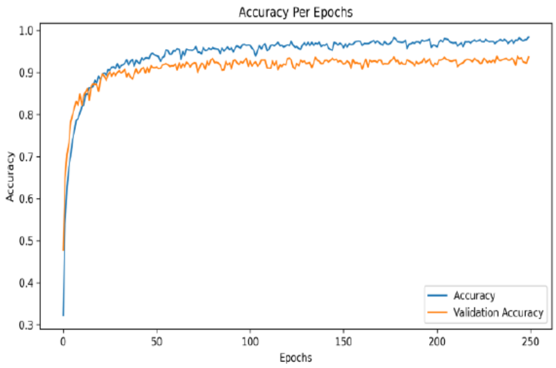

Fig. 8. Accuracy & Validation Accuracy graph

Loss – Shows how well the model fit to the training set of data. As the model learns from the data, it ought to get smaller over time.
Validation loss - Validation loss provides an indication of how well the model's fit to fresh data. The validation loss is a useful tool for evaluating the model's generalization ability to data that was not used in training.

The below figure shows that the loss is lesser than validation loss for each epochs resulting in efficient output.

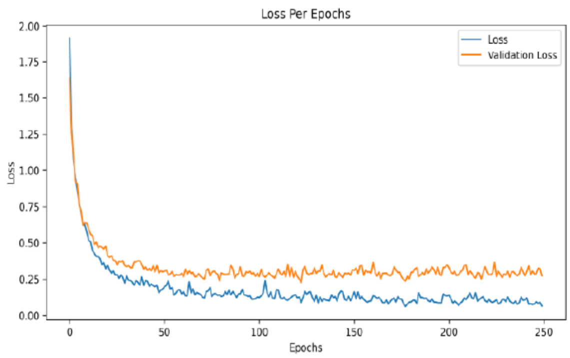

Fig.9 . Loss & Validation Loss graph

C. EMBEDDED ARDUINO OUTPUT

The indication of prediction of ambulance siren & other vehicle sounds is fed into the Arduino to dynamically adjusts the traffic signal to prioritize the way for emergency vehicle using the below code. After reading the input, if the value at the serial monitoring is "1" which is shown on the LCD, the traffic light LED is changes to GREEN if the result indicates that the ambulance siren has been detected; if not, the traffic signal operates at RED.

```
#define PIN_RED 2
#define PIN_YELLOW 3
#define PIN_GREEN 4

void setup() {
Serial.begin(9600);
pinMode(PIN_RED, OUTPUT);
pinMode(PIN_YELLOW, OUTPUT);
pinMode(PIN_GREEN, OUTPUT);
}

void loop() {
if(Serial.available()>0)
{
int a=Serial.read();
if(a=='1')
{
Serial.print("1");
digitalWrite(PIN_GREEN,LOW);
digitalWrite(PIN_RED,HIGH);
delay(3000);
}
else
{
Serial.print("0");
digitalWrite(PIN_GREEN,HIGH);
digitalWrite(PIN_RED,LOW);
delay(3000);
}}}
```

TABLE II . INDICATION OF TRAFFIC LED ACCORDING TO THE AUDIO RECOGNITION

Arduino software input	Output traffic signal
1 (i.e., Ambulance Siren)	Green
0 (i.e., Other traffic sounds)	Red

VII. CONCLUSION

The creation of a Tensor Flow-based "automatic recognition of ambulance siren system" that can instantly recognize ambulance siren in traffic signal has the potential to improve public safety and emergency response. The system can distinguish ambulance sirens from other noises with accuracy by utilizing machine learning techniques, which could enhance the effectiveness of emergency services. As well as this embedded system consisting of dynamic traffic signal adjustment improves emergency response and enhances emergency management and contributes to more efficient healthcare delivery in critical situation. This invention has the potential in saving lives of the people at risk. Additionally, this strategy will help intelligent traffic signals prioritize emergency vehicles and guarantee that they proceed smoothly through intersections.

REFERENCES

[1] S.Sathruhan, O. K. Herath, T. Sivakumar and A.Thibbotuwawa, "Emergency vehicle detection using vehicle sound classification: A deep learning approach," 2022 6th SLAAI international conference on artificial intelligence (SLAAI-ICAI), Colombo, Sri Lanka, 2022.

[2] Usaid, Muhammad & Rajab, Tabarka & Rashid, sheikh Muhammad & Hussain, Syeda. (2022). “Ambulance siren detection using artificial intelligence in urban scenarios.” Sir Syed research journal of engineering & technology, vol. 12, no. 1, 2022.

[3] Haque, s., Sharmin, S., Deb, K. (2022), “Emergency vehicle detection using deep convolutional neural network”, In: Uddin, M.S., Jamwal, P.K., Bansal, J.C. (Eds) proceedings of international joint conference on advances in computational intelligence. Algorithms for intelligent systems. Springer, Singapore.

[4] M. Gao, Z. Wang, Q. Liu, W. Xiong, and J. Wu, “Emergency vehicle identification based on convolutional neural networks and sound signals," IEEE access, vol. 8, pp. 158707-158716, sept. 2020.

[5] Becker, L. R., Zaloshnja, E., Levick, N., Li, G., & Miller, T. R. (2003). The relative risk of injury and death in ambulances and other emergency vehicles. Accident Analysis & Prevention, 35(6), 941- 948.

[6] Foggia, P., Petkov, N., Saggese, A., Strisciuglio, N., & Vento, M. (2015). Audio surveillance of roads: A system for detecting abnormal sounds. IEEE transactions on intelligent transportation systems, 17(1), 279-288.

[7] Liaw, J. J., Wang, W. S., Chu, H. C., Huang, M. S., & Lu, C. P. (2013, October). Recognition of the ambulance siren sound in Taiwan by the longest common subsequence. In 2013 IEEE International Conference on Systems, Man, and Cybernetics (pp. 3825-3828). IEEE.

[8] M. Gao, Z. Wang, Q. Liu, W. Xiong, and J. Wu, "Emergency Vehicle Identification Based on Convolutional Neural Networks and Sound Signals," IEEE Access, vol. 8, pp. 158707-158716, Sept. 2020.

[9] B.Li, C.Li, and Q.Liu, "Emergency Vehicle Detection Using Multi-Modality Sensor Data," Sensors, vol. 20, no. 23, pp. 6877, Nov. 2020.

[10] J.Zhang, X.Wang, and Z.Li, "Emergency Vehicle Identification Based on Convolutional Neural Network and Wavelet Transform," Sensors, vol. 21, no. 3, pp. 936, Jan. 2021.

[11] Liaw, J. J., Wang, W.S., Chu, H. C., Huang, M. S., & Lu, C. P. (2013, October). Recognition of the ambulance siren sound in Taiwan by the longest common subsequence. In 2013 IEEE International Conference on Systems, Man, and Cybernetics (pp. 3825-3828). IEEE.

[12] A.Abramson and I.Cohen, “Simultaneous Detection and Estimation Approach for Speech Enhancement,” in IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 8, pp. 2348–2359, Nov. 2007.

[13] B.Han and E.Hwang, “Environmental sound classification based on feature collaboration,” 2009 IEEE International Conference on Multimedia and Expo, New York, NY, 2009, pp. 542–545.

[14] B.D.Barkana and I.Saricicek, “Environmental Noise Source Classification Using Neural Networks,” 2010 Seventh International Conference on Information Technology: New Generations, Las Vegas, NV, 2010, pp. 259–263.

[15] G.Hinton et al., “Deep Neural Networks for Acoustic Modelling in Speech Recognition: The Shared Views of Four Research Groups,” in IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82–97, Nov. 2012.