Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 227–234

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

Acoustic Based Emergency Vehicle Detection Using Ensemble of deep Learning Models

Usha Mittal[1], Priyanka Chawla[2]*

[1]School of Computer Science Engineering, Lovely Professional University, Phagwara, Punjab
[2]National Institute of Technology, Warangal

Abstract

The temporal and spectral structure is possessed in the time-frequency domain by sound events. Analyzing and classifying acoustic environment using sound recording is an emerging research area. Convolutional layers can quickly extract high-level features and shift-invariant features from the time-frequency domain. In this work, emergency vehicle detection (EVD) like fire brigades, ambulances, and police cars is done based upon their siren sounds. Dataset from Google Audioset ontology was collected and features are extracted by Mel-frequency Cepstral Coefficient (MFCC). Three deep neural networks (DNN) models (dense layer, Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN)) with different configurations and parameters have been investigated. Then, an ensemble model has been designed with optimum selected models by performing experimental tests on various configurations with hyper-parameter tuning. The proposed ensemble model provides the highest accuracy of 98.7%, while the recurrent neural network (RNN) model provides an accuracy of 94.5%. Also, performance analysis of deep learning models is done with various machine learning models like Perceptron, SVM, decision tree etc.

Keywords: Audio recognition; CNN; DNN; emergency vehicle detection; MFCC; RNN; siren sound

1. Introduction

Special siren sound signals are used by emergency vehicles to identify them on roads. However, traffic jams or road emergencies can cause emergency services to be delayed. Thus, to avoid delay due to red signal at the traffic signal intersection, EVD has been focused on their siren sounds. Siren sounds have been separated into three categories: ambulance, fire truck, and police car. Each country has its regulations on the siren sound type and frequency band.

Generally, the sirens are warning signals issued in an emergency and standardized by the International Organization of Standard (ISO), and ISO 7731 [1] gives important guidelines for warning sirens. Audio recognition using an ensemble of deep learning models is the primary approach used in this paper. Before performing the recognition task, audio feature extraction methods are used to obtain valuable features in the time-domain and frequency- domain. Intelligent transport systems can use the application of the EVDs system. Traffic controllers can integrate siren detection to prioritize direction with emergency vehicles by changing the signal status and adjusting the timing of the green signal accordingly. The major contributions of this work are as follows:

*Corresponding author.
Email address: priyankachawla.cse@gmail.com (Priyanka Chawla)

1877-0509
10.1016/j.procs.2023.01.005

I. Data collection, pre-processing, and feature extraction from audio files downloaded from open source library.
II. Implementation of fully connected neural network (NN) and CNN models with different configurations consisting of different layers and parameters.
III. Implementation of RNN model and selection of appropriate hyper parameters by considering different configurations.
IV. Designing an ensemble model using three deep learning models, i.e., fully connected (FCNet), CNN model (CNN_Net), and RNN model (RNN_Net), for the classification of siren sounds

Paper organizing is as follows: In Section II, related work to acoustic-based emergency vehicle detection has been done. In Section III, different models have been investigated and analyzed for the classification of siren sounds. Then, experiment results have been given in Section IV, and Section V contains the conclusion.

2. Related Work

Studies show that research done in siren sound recognition is significantly less like [2– 13]. J. Liaw et al. suggested identifying the siren of ambulance using Longest Common Subsequence (LCS) [2] in Taiwan. The proposed model provided an accuracy of 85%. Mel- Frequency Cepstral Coefficient (MFCC) based speech recognition technology was introduced in [3], with multilayer neural networks through the majority voting techniques to detect the siren sounds. The model defined in [3] had low computational complexity. Still, it could not provide efficient analysis on noisy and diverse datasets, and a reproduction technique was used to increase the training and testing data. In [4], sirens were detected with two different methods, i.e., a multilayer neural network (MNN) and a sinusoidal model system. MNN system was taken from the speech recognition. The sinusoidal model system utilized siren sounds and tried to extract signals from the background noises to minimize noise interference. Both the methods were evaluated on a small dataset, and both the model gave almost similar accuracy. In [5], part-based models (PBMs) were proposed by the author, which was initially used in computer vision to detect sirens in noisy traffic considering spectro-temporal domain. When trained on MFCC or log-mel attributes, PBMs performed better than hidden Markov models (HMMs). However, its performance was below 90%.

A two-stage detector for audio-based detection was proposed by L. Marchegiani et al. [6] in intelligent vehicles. The initial stage was used to detect an irregular sound, and the second was responsible for removing or minimizing noise and classification. The idea in [6] had been taken from image processing. Each incoming signal's spectrogram is treated as an image and a segmentation method was used to remove and distinguish the target signal. The KNN (K-Nearest Neighbor) algorithm was used after the noise was removed and provided an accuracy of 83%. In [7], siren detection was performed by analyzing audio signals based upon digital signal processing methods, like estimating frequency components in a specified frequency range. In [8], SVM with feature selection techniques was used to detect alarm sounds. It provided an accuracy of more than 90% on a small dataset. The major limitation of this work was the time consumed in feature engineering.

Various works have been done using microcontrollers [9, 10, and 11] and hardware de- sign [12, 13] to detect alarm sounds. An ambulance was detected considering siren sound in [9] by performing Fast Fourier Transform (FFT) twice on a microcontroller. Though the proposed method could work under the Doppler Effect, it is not the preferred choice due to the high computational cost. A micro-controller-based system using frequency and siren sound periodic repletion characteristics were used by F. Meucci et al. [10] for detecting emergency vehicles. The major limitation of this model is that it was evaluated only for the type of sirens with the frequency of 392 Hz and 660 Hz. Durbin's recursive method was used for hard-of-hearing drivers with the help of a linear prediction model [11]. R. Dobre et al. [12, 13] designed an analog electronics circuit-based system with low computational cost. The SPICE simulator was used to evaluate the siren signal. It gave acceptable accuracy on small datasets but failed to provide performance on a large dataset. Fatimah et al. [25] suggested a model for emergency vehicle detection in which signals were processed using bandpass filters. Two types of features were used and different ML models like KNN, SVM and ensemble were compared for selection of best model.

CNN is widely used and famous for audio recognition applications like music tagging, automatic speech recognition (ASR) [18, 19], and environmental sound classification appli- cations [14–17]. For environmental sound classification, GoogleNet and Alexnet are the two most popular image recognition networks explored by V. Boddapati et al. [14]. In these net- works, spectrogram and MFCC were given input data for training the model. The proposed model

yielded an accuracy of up to 90%, demonstrating the [14] approach's potential. K. Piczak [16] and J. Salamon et al. [15] proposed CNN-based models for environmental sound classification. Considering log-mel spectrogram data for training, models presented in [15] and [16] provided almost equal accuracy less than 80%. Baghel et al. [26] utilized YOLO model with two phases for recognition of emergency vehicles. One phase was for the generation of bounding boxes and other for classification. Tran et al. [27] suggested audio and vision based model for recognising emergency vehicles in which YOLO was utilized for image processing and WaveResNet was utilized for audio processing.

The significant limitations of the work carried till now for EVD are (1) the lack of experimental data, (2) use of hardware-based systems and shallow algorithms (3) use of handcrafted features from the time domain or frequency domain for the training of models. Considering all, EVD has been improved by collecting the dataset from an open-source library, i.e., Audio set ontology, and employing deep learning models such as CNN and RNN.

3. Proposed Methodology

In this work, recurrent neural network and an ensemble of three different deep learning models has been created and evaluated to classify siren sounds of emergency vehicles. Base estimators of proposed ensemble consist of fully connected neural network, convolution neural network and recurrent neural network. The details of the base estimators are as follows:

- *Fully Connected NN (FCNet)*: This architecture is purely based on dense layers without any convolutional layer. This network is evaluated with the different number of fully connected layers up to 8 with the various parameters for selection of best model.
- *Convolutional NN (CNN_Net)*: This architecture has a variable number of 2D convolutional layers up to 6 and variable number of filters, and a 4X4 kernel size. After convolutional layers, the max-pooling layer is used to prevent overfitting. Further, the dropout layer with a 0.25 parameter is applied after the dense layer.
- *Recurrent NN (RNN_Net)*: This architecture is a recurrent neural network (RNN) that consists of different number of long short-term memory (LSTM) layers with a different number of neurons.

Ensemble Model: After evaluating and investigating various experimental results of different configurations of three models, an ensemble has been designed, making the prediction based upon majority voting. The base network of FCNet contains eight layers, CNN_Net consists of 3 layers, and RNN_Net comprises five layers. The proposed ensemble network's architecture is shown in Figure 1. The working of the proposed system in also explained in the algorithm 1.

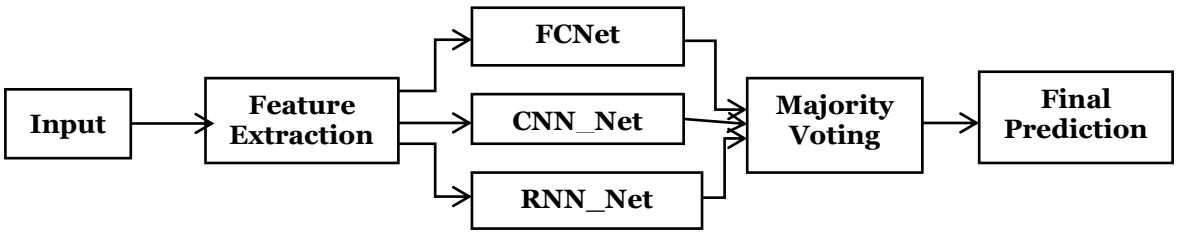

Figure 1: Architecture of Proposed Ensemble Model

Algorithm 1: Proposed Methodology

Input: An audio file
Output: The predicted class of emergency vehicle.
Begin:

I. Extract features from the given audio file using MFCC technique.
II. II. Provide extracted features to three base models i.e. FCNet, CNN_Net and RNN_Net and store their predictions: y_FCNet, y_CNN_Net, y_RNN_Net respectively.
III. III. Apply majority voting on the obtained predictions: mode(y_FCC_Net, y_CNN_Net, y_RNN_Net) and return final prediction

End

4. Experimental Results and Discussion

a. Data Collection

The experimental dataset is collected from Google Audioset Ontology [24] which contains sound events in a hierarchal arrangement. This ontology consists of various sounds such as animal, human, environmental, musical, and miscellaneous. It includes the siren sounds of four different types of vehicles, i.e., Police Car, Ambulance, Fire Engine, and Civil Defence Siren, in video format. Information about the video is available in a CSV file which contains the YouTube link to a video, starting time, ending time, and label. From the dataset, the video of three types of emergency vehicles (Ambulance, Police car, Fire truck) has been downloaded by using two python libraries, i.e., "pafy" and "youtube_dl." The whole audio file is of no use, so the siren sound of the vehicle is clipped from downloaded files using the "moviepy" library.

b. Feature Extraction

Although many feature extraction methods are available to extract features from audio data, Mel Frequency Cepstral Coefficient (MFCC) is used in this work. It has extracted 39 different features from the dataset where the first feature corresponds to the audio pitch, and 12 of them are related to the amplitude of frequencies. The flow chart of the features ex- traction is given in figure2. To extract useful information from audio files, the "Librosa" [20] library is used. The final shape of the feature vector and target is (259169, 40) and (1,301), respectively. Figure 3 shows the waveform of the audio of police car siren sound.

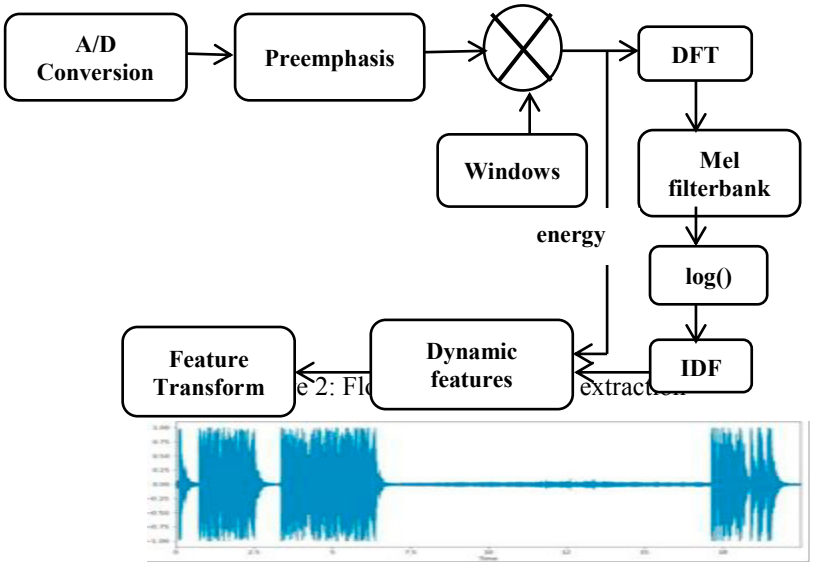

Figure 3: Waveform of police car siren

c. Hyperparameter Tuning

In this work, three types of deep learning models have been investigated and analysed.
When dealing with deep learning models, it is necessary to select the appropriate number of layers and parameters for the network to optimize its performance. Thus, a series of experiments have been conducted. The impact of the various layers and parameters is investigated on all three models. Following that, appropriate models from each configuration to acoustic-based EVD have been selected. Tables 1, 2, and 3 shows the training and testing accuracy on different layers and parameters used in FCNet, CNN_Net, and RNN_Net models, respectively.

For implementing deep learning architectures, "TensorFlow" has been used. It is a free and open-source library for mathematically extensive programming mainly focusing on ma- chine learning and neural networks, developed by Google. All the models have been trained on Google Colaboratory, supporting GPU (Graphics Processing Unit) free for public use. In all the configurations, the Relu activation function is used at the hidden layer, and at the output

layer, Softmax activation is applied. The models are trained using the Adam optimizer [22] with a learning rate of 0.001 and decay of 0.0001, and categorical cross-entropy loss.

Table 1: Layers and parameters in multilayer fully connected neural network with different number of fully connected layers

Layer	FC Layer-2	FC Layer-3	FC Layer-4	FC Layer-5	FC Layer-6	FC Layer-7	FC Layer-8
Input	0	0	0	0	0	0	**0**
FC-1024	35267584	35267584	35267584	35267584	35267584	35267584	**35267584**
FC-512	524800	524800	524800	524800	524800	524800	**524800**
FC-512	262656	262656	262656	262656	262656	262656	**262656**
FC-512	—-	262656	262656	262656	262656	262656	**262656**
FC-256	—-	—-	131328	131328	131328	131328	**131328**
FC-256	—-	—-	—-	65792	65792	65792	**65792**
FC-128	—-	—-	—-	—-	32896	32896	**32896**
FC-64	—-	—-	—-	—-	—-	8256	**8256**
FC-32	—-	—-	—-	—-	—-	—-	**2080**
Output-3	1539	1539	771	771	387	195	**99**
Total Parameters	36,056,579	36,319,235	36,449,795	36,515,587	36,548,099	36,556,163	**36,558,147**
Training Accuracy %	100	99.58	100	99.58	100	100	**100**
Testing Accuracy %	60	70	75	84	92	94.6	**96.4**

Table 2: Layers and parameters in a convolutional neural network with different number of 2D convolutional layers

Layer	Conv_Layers-2	Conv_Layers-3	Conv_Layers-4	Conv_Layers-5	Conv_Layers-6
Input	0	**0**	0	0	0
Conv 4X4 – 32	544	**544**	544	544	544
Conv 4X4 – 32	16416	**16416**	16416	16416	16416
Conv 4X4 – 64	—-	**32832**	32832	32832	32832
Conv 4X4 –64	—-	**—-**	65600	65600	65600
Conv 4X4 – 128	—-	**—-**	—-	131200	131200
Conv 4X4 – 128	—-	**—-**	—-	—-	262272
FC – 512	564,265,472	**281805312**	70451712	35062272	6947328
FC- 64	32832	**16416**	32832	32832	32832
Output 3	195	**99**	195	195	195
Total	564315459	**281871619**	70600131	35341891	7489219
Training Accuracy %	100	**99.58**	93.3	95	95
Testing Accuracy %	61	**92.4**	85.3	88.6	84.4

Table 3: Layers and parameters in anRNN with different long short term memory (LSTM) layers

Layer	LSTM_Layer-2	LSTM_Layer-3	LSTM_Layer-4	LSTM_Layer-5	LSTM_Layer-6
Input	0	0	0	**0**	0
LSTM 32	9344	9344	9344	**9344**	9344
LSTM 32	8320	8320	8320	**8320**	8320
LSTM 64	—-	24832	24830	**24832**	24832
LSTM 64	—-	—-	33024	**33024**	33024
LSTM 128	—-	—-	—-	**98816**	98816

LSTM 128	—-	—-	—-	—-	131584
FC 128	4224	8320	8320	**16512**	16512
Output 3	387	387	387	**387**	387
Total	22275	51203	84227	**191235**	322819
Training Accuracy %	84.07	89.6	92.2	**98.7**	90.4
Testing Accuracy %	61.29	75.7	85.2	**94.5**	84.1

d. Results and Discussions

Comparative analysis of the proposed model with different deep learning architectures is given in figure 4. In this paper, four different models have been explored. FC_Net model which consists of only dense layers provide accuracy of 96.4% and its inference time is 0.061s. CNN_Net model provides an accuracy of 92.4% while RNN_Net and Ensemble have accuracy of 94.5% and 98.7% respectively. A comparison based upon inference time is also given in table 4 which clearly shows that time taken by RNN_Net and FCNet is almost same, while CNN takes longer time than these models and response time of Ensemble is the highest and it takes almost 1.5seconds.

Various machine learning models are also evaluated on the collected dataset and results are compared with the deep learning models. Table 5 shows the comparisons of different machine learning models with proposed deep learning models. Although decision tree and random forest provides higher training accuracy but their testing accuracy is very low and models are over-fitted. Compared to machine learning models, proposed deep learning models provide better accuracy and acceptable models.

Several kinds of research based on microcontrollers [9–11] and circuit design [12, 13] only stated the prospects of the siren detection system without evaluating its accuracy on a large dataset. As a result, works focused on machine learning (ML), and deep learning (DL) methods have been considered for comparison. Table 6 compares our proposed models to previous findings [2, 6, 8, 21, and 23] in terms of methods and functionality, and prediction accuracy. Table 6 depicts that classification accuracy of CNN model proposed by L. Marchegiani [21] and proposed model with RNN is almost same. Machine learning models like KNN [6], HMM [8], and part based models [8] had accuracy less than 90%. CNN mod- els developed by Tran [23] achieved accuracy up to 98.24%. The proposed ensemble model provides 98.7% accuracy which is highest among all works. The proposed ensemble yields promising results 98.7% which is better than results of proposed RNN_Net (94.5%) and other related works [2] 85%, [6] 83%, , [8] 86%, [21] 94%, [23] 98.24%. The performance of the ensemble model is higher than the results of the remaining models. A comparative analysis of the proposed model is given in Table 6.

Table 4: Comparative Analysis of Different Models

Model	Accuracy	Inference Time (s)
FCNet	96.4	0.061
CNN_Net	92.4	0.151
RNN_Net	94.5	0.061
Ensemble	98.7	1.5

Table 5: Comparative Analysis of deep learning models with machine learning models

Method	Training Accuracy	Testing Accuracy
Perceptron (L1 Regularization)	59.5	49.2
Logistic Regression (L2 Regularization)	65.8	44.3
SVM (Kernel=Polynomial)	61-7	52.4
KNN (neighbors=10)	61.7	46
Decision Tree (Entropy)	100	57.3
Random Forest (estimators=12)	98	57.3
Naïve Bayes	65	42.6
ADABoost Classifier (base=Naïve Bayes, number of estimators=11)	51	51

Fully Connected Neural Network (8 FC layers)	100	96.4
Convolutional Neural Network (3 conv layers)	99.58	92.4
Recurrent Neural Network (5 LSTM layers) (Proposed)	98.7	94.5
Ensemble Model (Proposed)	99.74	98.7

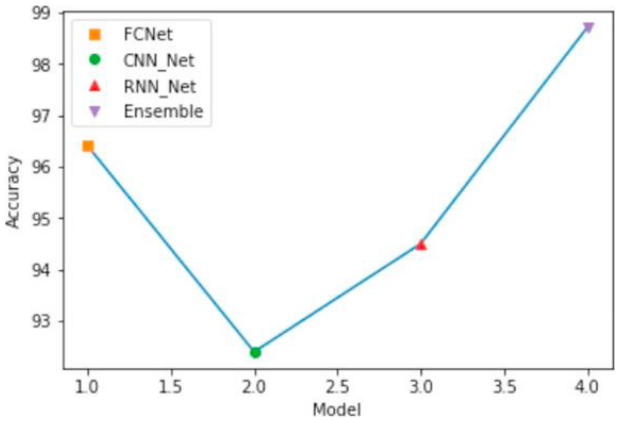

Figure 4: Comparison of various DL models based on Accuracy

Table 6: Comparative Analysis of Proposed model with Existing Methods

Model/Technique	**Accuracy(%)**
[2]	85
[6]	83
[8]	86
[21]	94
[23]	98.24
RNN_Net (Proposed)	94.5
Ensemble (Proposed)	98.7

5. Conclusion and Future Work

This paper introduces an ensemble of deep learning-based models for siren sound-based emergency vehicle detection. The proposed model consists of the fully connected model, CNN model, and RNN model. Models have been trained on MFCC features extracted from collected data. Various experimental results have been obtained, and they demonstrate that the proposed model is more efficient than other existing models. The proposed ensemble model provides the highest accuracy of 98.7%, while the recurrent neural network (RNN) model provides an accuracy of 94.5%. Also, performance analysis of deep learning models is done with various machine learning models like Perceptron, SVM, decision tree etc. Proposed model can be implemented in real time for detecting emergency vehicles approaching towards the intersection and priority can be provided to reduce their waiting time.

Acoustic-based models are also a better choice as compared to image-based detection models. As emergency models move at high speed and it is difficult to capture the image of the emergency vehicle using the cameras. But emergency vehicles start giving warnings from long distances that the system can easily capture and process.

Even though the proposed ensemble produces satisfactory results, more work is required to improve detection efficiency and meet the criteria for a reliable and convenient emergency vehicle detection system. Furthermore, work could be done to localize siren sounds so that direction of emergency vehicles can be identified.

References

[1.] "ISO 7731: Ergonomics -Danger signals are further subdivided and work areas -Auditory danger signals," International Organization for Standardization, 2013.

[2.] J.J. Liaw, W.S. Wang, H.C. Chu, M.S. Huang and C.P. Lu, "Recognition of the ambulance siren sound in Taiwan by the Longest Common Subsequence," IEEE Int. Conf. on Systems, Man, and Cybernetics, 2013.

[3.] F. Beritelli, S. Casale, A. Russo, and S. Serrano, "An automatic emergency signal recognition system for the hearing impaired," In Proceedings of 12th Digital Signal Processing Workshop and 4th Signal Processing Education Workshop, Sept 2006, pp. 179-182

[4.] Daniel P.W. Ellis, "Detecting alarm sounds," in Proceedings of the Recognition of real-world sounds: Workshop on consistent and reliable acoustic cues, Aalborg, Denmark, 2001, pp. 59–62.

[5.] J. Schroder, S. Goetze, V. Grutzmacher, Jorn Anem uller, "Automatic acoustic siren detection in traffic noise by Part-based Models," IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), 2013.

[6.] L. Marchegiani, I. Posner, "Leveraging the urban soundscape: Auditory perception for smart vehicles," IEEE Int. Conf. on Robotics and Automation (ICRA), Singapore, 2017, pp. 6547-6554.

[7.] Ondrej Karpis, "System for Vehicles Classification and Emergency Vehicles Detection," IFAC Proceedings Volumes, Volume 45, Issue 7, 2012, pp. 186-190

[8.] D. Carmel, A. Yeshurun and Y. Moshe, "Detection of alarm sounds in noisy environments," 25th European Signal Processing Conference (EUSIPCO), Kos, 2017, pp. 1839-1843.

[9.] T. Miyazaki, Y. Kitazono, M. Shimakawa, "Siren detector using FFT on dsPIC," in Proceedings of the 1st IEEE/IIAE Int. Conf. on Intelligent Systems and Image Processing, 2013, pp.266-269.

[10.] F. Meucci, L. Pierucci, E. Del Re, L. Lastrucci, P. Desii, "A realtime siren detector to improve safety of guide in traffic environment," 16th European Signal Processing Conference (EUSIPCO 2008), pp.25-29.

[11.] S.W. Park and J. Trevino, "Automatic detection of emergency vehicles for hearing impaired drivers," Texas A&M UniversityKingsville, EE/CS Department, MSC 192, Kingsville, TX 78363.

[12.] R. A. Dobre, V. A. Niţă, A. Ciobanu, C. Negrescu and D. Stanomir, "Low computational method for siren detection," IEEE 21st Int. Symposium for Design and Technology in Electronic Packaging (SIITME), Brasov, 2015, pp. 291-295.

[13.] R. A. Dobre, C. Negrescu and D. Stanomir, "Improved low computational method for siren detection," IEEE 23rd Int. Symposium for Design and Technology in Electronic Packaging (SIITME), Constanta, 2017, pp. 318-323.

[14.] V. Boddapati, A. Petef, J. Rasmusson, L. Lundberg, "Classifying environmental sounds using image recognition networks," Procedia Computer Science, Volume 112, 2017, pp. 2048-2056.

[15.] J. Salamon and J.P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," IEEE Signal Processing Letters, Nov 2016.

[16.] K.J. Piczak, "Environmental sound classification with convolutional neural networks," Int. Workshop on Machine Learning for Signal Processing, Boston, USA, Sep.2015, pp.17-20.

[17.] J. Lee, T. Kim, J. Park, and J. Nam. "Raw Waveform-based Audio Classification Using Sample-level CNN Architectures," arXiv:1712.00866v1 [cs.SD] 4 Dec 2017.

[18.] S. Thomas, S. Ganapathy, G. Saon and H. Soltau, "Analyzing convolutional neural networks for speech activity detection in mismatched acoustic conditions," in Proc. IEEE ICASSP, Florence, 2014, pp. 2519-2523.

[19.] O. Abdel-Hamid et al., "Convolutional Neural Networks for Speech Recognition," in IEEE/ACM Trans. Audio, Speech, Language Process., vol. 22, no. 10, pp. 1533-1545, Oct. 2014.

[20.] B. McFee, C. Raffel, D. Liang, D.P.W. Ellis, M. McVicar, E. Battenberg, and O. Nieto. "librosa: Audio and music signal analysis in python," in Proceedings of the 14th python in science conference, pp. 18-25. 2015

[21.] L. Marchegiani, and P. Newman, "Listening for Sirens: Locating and Classifying Acoustic Alarms in City Scenes," ArXiv, abs/1810.04989, 2018.

[22.] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Int. Conf. on Learning Representations, 2014, arXiv:1412.6980.

[23.] TRAN, V., & TSAI, W. (2020). Acoustic-based Emergency Vehicle Detection Using Convolutional Neural Networks. IEEE Access. https://doi.org/10.1109/ACCESS.2020.2988986

[24.] AudioSet. Retrieved from https://research.google.com/audioset/dataset/index.html Dataset

[25.] 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 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Jul. 2020, pp. 1–6, doi: 10.1109/ICCCNT49239.2020.9225414.

[26.] 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 Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019). Lecture Notes in Networks and Systems, 2020, vol. 121, pp. 607–618, doi: 10.1007/978-981-15-3369-3_45.

[27.] V.-T. Tran and W.-H. Tsai, "Audio-Vision Emergency Vehicle Detection," IEEE Sensors Journal, vol. 21, no. 24, pp. 27905–27917, Dec. 2021, doi: 10.1109/JSEN.2021.3127893.