

Ambulance Siren Detection using ANN

R.Kamaladevi¹, M.Mohamed Hashir² and Y.Godbin James³

¹Assistant Professor, Mepco Schlenk Engineering College, Anna University, Chennai, India.

Email: kamalachozhan@mepcoeng.ac.in

²⁻³Under Graduate students, Mepco Schlenk Engineering College, Anna University, Chennai, India.

Email: {mohamed.hashir67, godbinjames02}@gmail.com

Abstract—Hearing and seeing accidents and deaths happening all over the world due to poor road networking and overriding of traffic rules is just tragic. And these accidents involve an ambulance rushing to nearby hospital in the hope to save the individual's life, and what they undergo on the way is that the congestion of traffic signals. Now the major setback found in this unhealthy situation is that the traffic police couldn't identify the direction of ambulance with only its siren, and they could only intervene and change the light signals only when they see the ambulance. This unawareness of ambulance's direction gave an opportunity to propose a solution to solve the problem by assisting in existing traffic light concept with the modification in its generic system by making use of the ambulance's siren feature. Machine Learning is now permeable in every product or solution that exists, and it really does its job on helping us and it has also given us a way to solve the existing traffic problem during emergency situations making the traffic light system automatically analyse the arrival of Emergency Vehicles (EV) and assist the traffic police beforehand on which direction the ambulance is arriving so that emergency measures like changing the lights to avoid congestion in the intersection can be done. By implementing Artificial Neural Networks (ANN) with Mel-Frequency Cepstral Coefficients (MFCC) feature extraction method the model achieved a 90% accuracy and accurately predicted the siren sound which was then verified by the CSV file given by the dataset-Urbansound8k. The audio data were converted into data points using MFCC and then easily mapped them with their respective classes. With ANN and MFCC, the prediction and accuracy rate are said to be better and by simple neural networks high grade feature extraction method was implemented.

Index Terms— Machine learning, Traffic signal, MFCC, Ambulance, ANN.

I. INTRODUCTION

One in ten people die on the way to hospital because of the delay of ambulance in reaching the hospital due the heavy traffic and one of the major reasons for that is the ambulance siren can't reach long distance to alert other vehicles and the traffic police to make the way for the ambulance, alert can be made only when the ambulance and its direction of motion is detected. The ambulance siren is a great feature that is being used not to its full potential. The primary function of the siren in the ambulance is to express the urgency of the vehicle to the drivers on road. The emergency varies from one level to another so sirens are varied by the intensity and the loudness according to the situation. It can be realized that the drivers on-road can only hear the ambulance incoming when the ambulance is near them. Most of the times this causes unethical competition between drivers and leads to collision in the traffic junctions. In this paper the ambulance siren is detected from the audio input

by better audio-feature extraction method using deep learning methods. Taking it to the next level, the future work can be proposed as programming with embedded devices to produce alert to the traffic police and to other vehicles by using speakers in the road junctions and the colour of the traffic lights can be changed based on the direction of the ambulance so that the way for the ambulance is cleared even before the ambulance reaches the junction. This can reduce the delay of the ambulance to the greater extent such that the valuable lives of the patient can be saved. To carry out the feature extraction of siren process an emerging method called Mel-Frequency Cepstral Coefficients (MFCC) is introduced. The basics and workings of MFCC are elaborated in [3] and [10]. Basically, MFCC in architectures that need high accurate sampling data points which can be efficiently reproduced as array of data that helps in understanding the pitch and frequency of the wave in particular point. Now, by utilizing this MFCC feature to siren audio data combining with high architecture neural networks would be useful in detection of whether an ambulance is approaching on its way towards the traffic junction or not. Regarding the usage of high neural networks with MFCC is elaborated in [10]. But the higher-class models can't be relied on scalability when it comes to adding new computational features to the system. And also, to detect the siren sound there is no need of custom and more computational machine learning models. So, this proposed idea focusses on achieving the same accuracy using simple Artificial Neural Network (ANN) model. And the purpose of this simple ANN model for siren detection is to detect ambulance with the help of Machine learning model and how this simple model can be used in IoT applications and can provide with scalable solution.

II. LITERATURE SURVEY

There have been significant papers proposing solutions for the existing traffic congestion during emergencies. In [1], the author Stefano proposed a ML solution where the ml model learns with siren's sound in baby steps and classify the ambulance sound in real-time. An Emergency Vehicle Priority System (EVPS) is proposed in [2] where based on type and severity of the accident the signal interventions are intimidated. Rubens Cruz Gatto used real-time audio source with traffic noise to train his ML model which is proposed elaborately in [3]. An audio and vision-based EV detection system that can detect Emergency Vehicles (EV) detects and alert the drivers about the emergency up ahead. It has been proposed in [4] with YOLO-EVD algorithm by an accuracy of 95.5% and they also proposed a Wave-Res Net, an end-to-end convolutional neural network with accuracy of 98%. In [5], Extreme Learning Machines (ELM) which uses raw audio data to train it in online and decide with classification in a simple manner. Author Van-Linh Nguyen proposed a model in [6] where the time and severity of the accident is noted, and list of paths are determined for that ambulance to go to the nearest hospital with the traffic congestion in the intersections and a fastest driving path is given and the traffic signal in that path is selected and pre-empted with necessary light control status. In [8], the author has proposed a concept where vehicles are monitored by RGB camera to detect the direction of the vehicle passing and audio is captured to sync with it to see if the vehicle is a public or private transport. In [9], the author proposed a system where the drivers send their trip information to the traffic police in the traffic junction so that the system can optimize the traffic lights using Dijkstra's algorithm using SUMO and MATLAB server. In [10], the author used the MFCC audio features extracted from the audio input and developed a ML model with Bayesian Regularized Artificial Neural Network (BRANN). On all these papers the papers are constructed with audio processing techniques with complex platforms and tasks. As overcoming the complexes of audio processing techniques, this paper gives concept about using simple platforms to train and perform high-level models and providing insight on how this model can be beneficial to IoT system.

III. SOFTWARE

A. Dataset

A ML model is created with deep learning architecture of Artificial Neural Network (ANN) to classify the ambulance's siren sound so that the traffic intersection can be intimidated by the arrival of the ambulance with a confirmation. For this deep learning model, a dataset with siren audio and other audio which can possibly contribute to the noise in real-time is used to train the model. The dataset is 'Urbansound8k', which has 8732 audio files with the mix of audio classes. There are 10 different classes in the dataset, and they are children playing, air conditioner, car horn, dog bark, drilling, gunshot, jackhammer, siren, engine idling, street music. The dataset is balanced to 1000 files in all classes almost except siren has 929 files, car horn has 429 files and gunshot has 374 files represented in Table I. And additionally, 200 files of fire truck sounds are also added to differentiate ambulance siren and fire truck siren. And by that a total of 8932 wav audio files were used to train

the ML model. All the files have a slice filename to match it with its respective class. The dataset named ‘UrbanSound8k’ can be downloaded from its official website [7].

TABLE I. URBANSOUND8K DATASET SAMPLES

Class	No. of Samples
Dog bark	1000
Children playing	1000
Air conditioner	1000
Street music	1000
Engine idling	1000
Jack hammer	1000
Drilling	1000
Siren	929
Car horn	429
Gun shot	374
Fire truck	200

B. Google colab

Google colab is a platform where the python codes can be run for easy access and is especially suited for Machine Learning, data analytics and higher computation codes. Colab has a faster speed than PC’s CPU so that more complex computational and mathematical python codes can be executed without being interrupted. It has GPU services to make it even more iterative and faster. Additionally, Google drive can be integrated with google colab to connect datasets or importing libraries and models. Creating a ML model with google colab is very much easy compared to other platforms.

IV. PROPOSED IDEA

As ambulance uses its siren sound to let know the drivers in the road that an emergency has occurred, and the emergency vehicle needs way to move ahead but the real problem arises in every traffic junction where even though ambulance has arrived with its emergency siren “on” the drivers can only hear the siren sound when they are near them. The siren is efficiently used when the drivers know that the ambulance is arriving before they even see it so that they have sufficient time to move their vehicle and give way for the ambulance. So, the audio detection system is proposed to detect the siren sound with Machine learning models. Though, the high computational neural networks are not used in here to simplify the task of audio detection. The neural network like CNN is not used as it may have greater architecture but if it must be implemented on an IoT system the other embedded device may have compatibility and scalability issues. By keeping in mind that this detection system may be implemented in IoT system the need for simple computational neural network is needed. So, the proposed model has an ANN architecture with three layers as in input layer, hidden layer, output layer. The MFCC feature extraction method in audio data is the best way to analyse the audio data of its pitch and frequency with data points which is shown in Figure 1. These data points help in mapping the audio data on the areas of its high pitch and low pitch which can be efficiently used for matching an audio set. Combining these two, an ANN model by extraction of audio features using MFCC is carried out. In this way, a simple model but effective matching audio set can be created. The learning and accuracy of the model depends on how well the feature extraction is done. If the feature extracted has more data points, then it can be easily learned by ANN model thus creating a low computational machine learning model with greater feature extraction methods.

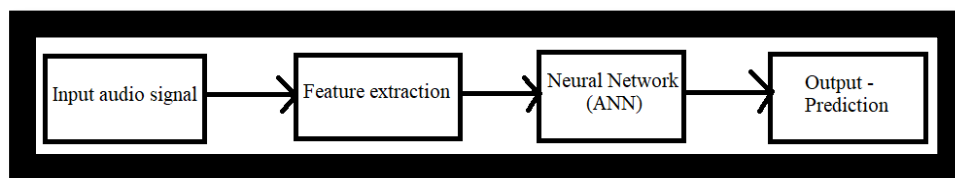


Figure 1. Block diagram of proposed model

A. ANN

Artificial Neural Network are inspired by biological neural networks which also works in it way by transferring the knowledge learnt from the samples from one neuron to the other and to other layers that are embedded with it as shown as in Figure 2. They are considered a traditional Machine learning model as it is a simple structure

model that can be easily used for any kind of learning and with that as a general structure, modifications can be made to create a custom and high computational neural network too.

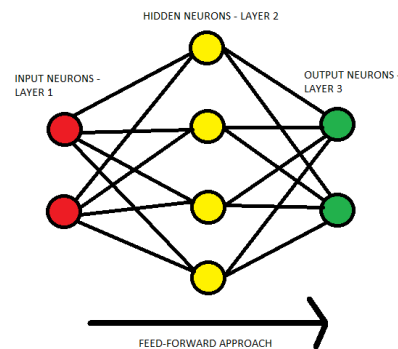


Figure 2. ANN layer model

B. Mel-frequency cepstrum

In sound processing, the mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFCC. Figure 3 gives the representation of a sample audio signal that is to be feature extracted using MFCC.

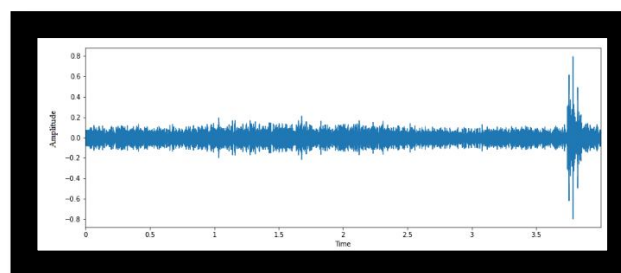


Figure 3. Sample audio signal for feature extraction

	feature	class
0	[-214.95764, 70.502464, -130.70279, -53.116936...	dog_bark
1	[-423.7311, 109.2299, -52.872597, 60.827538, 0...	children_playing
2	[-458.52844, 121.35432, -46.535675, 51.969467, ...	children_playing
3	[-413.63254, 101.61351, -35.43868, 53.047146, ...	children_playing
4	[-446.38693, 113.68634, -52.4572, 60.349724, 2...	children_playing

Figure 4. Feature extracted table

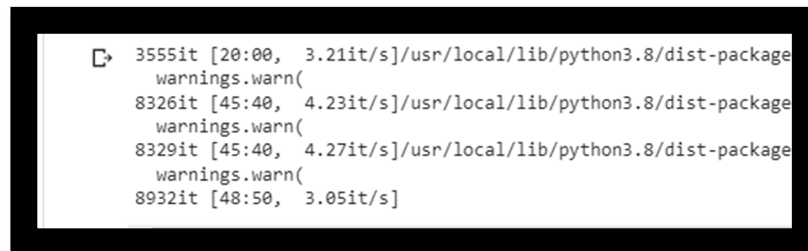
From Fig 4 it can be inferred that using MFCC the feature extraction of the audio data is implemented. The audio data was processed using Librosa library which is specifically used for audio pre-processing techniques. In librosa, there is MFCC functions that we can access. Now, due to the usage of librosa the sampling rate of the audio data is converted and normalized from 44.1KHz which is a standard sampling rate to 22.05 KHz. By converting and normalizing the sample rate there are no different samples in the audio and any mono or stereo based audio can be feature extracted by librosa using normalization. By this sampling rate, the audio data are pre-processed to extract the MFCC features, and those features are presented in an array of data points shown as in Figure 4. If inferred closely, not only features are extracted but its respective classes are also mapped. And all those same class audios have mostly similar data points after extraction. With this data points we can map the audio with its class.

C. Importance of Feature Extraction

- Feature extraction is used to:
- Reduce input data to neural network.
- Reduce redundancy present in input data.
- Provides an accurate representation of input data.

V. RESULTS AND DISCUSSIONS

The ML model and the prediction process is carried out in google collab to create and test the model of its accuracy.



```
3555it [20:00, 3.21it/s]/usr/local/lib/python3.8/dist-package
warnings.warn(
8326it [45:40, 4.23it/s]/usr/local/lib/python3.8/dist-package
warnings.warn(
8329it [45:40, 4.27it/s]/usr/local/lib/python3.8/dist-package
warnings.warn(
8932it [48:50, 3.05it/s]
```

Figure 5. Iteration rate

As there are 8732 files in urban-sound dataset [7] with different classes which also is applicable to use in siren detection as the siren is going to be detected among the noise in traffic signal. And relevant noises are already present as dataset in the given set. And additionally, 200 files of fire truck sounds are also added to differentiate ambulance siren and fire truck siren. And by that with a total of 8932 files the MFCC features are extracted in all the audio data thus going through 8932 iterations as shown in Figure 5. If the colab is changed to its runtime with GPU, the iterations can be done faster and if it's in CPU processing then it takes some time to iterate through all the data and extracting and storing them.



```
Epoch 100/100
216/224 [=====>...] - ETA: 0s - loss: 0.4269 - 
Epoch 100: val_loss did not improve from 0.32295
224/224 [=====] - 1s 4ms/step - loss: 0.4276
Training completed in time: 0:01:53.600612

test_accuracy=model.evaluate(X_test,y_test,verbose=0)
print(test_accuracy[1])

0.9015108942985535

from tensorflow.keras.models import load_model
model.save('/content/drive/MyDrive/ML Model/Audio_model_90.h5')
```

Figure 6. Accuracy of the model

From the above Figure 6 it is inferred that the ANN model implemented for 100 epochs and created, has an accuracy rate of 90%. To detect whether a given audio set is a 'siren' or not, the resultant accuracy is better. As the sound pitch of the ambulance siren are going to be high other than the noise created in the traffic roads with this accuracy siren sound can be detected easily and it is optimal enough to differentiate with other sounds.

Figure 7 represents the prediction of the siren sound with the developed ML model. The file 105289-0-1.wav is a siren sound and it has been extracted with features and passed through loaded ML model which in-turn predicted correctly with its class. The MFCC features and shape of the wav audio file is also displayed.

The CSV file for the dataset contains information about the filename, file ID, start and end of the audio signal, folder the audio is present, the class-ID for all respective audio and class of the audio. With this information, the audio is mapped for obtaining the audio and feature extracting and after the prediction process is completed, the predicted output can be confirmed with this dataset of its class and if needed class-ID, filename and folder the specific audio is present. As per Table II, the siren sound predicted has been verified with the CSV file created for the dataset. The class of the siren is 8 according to the CSV file and through trained model it has been accurately predicted.

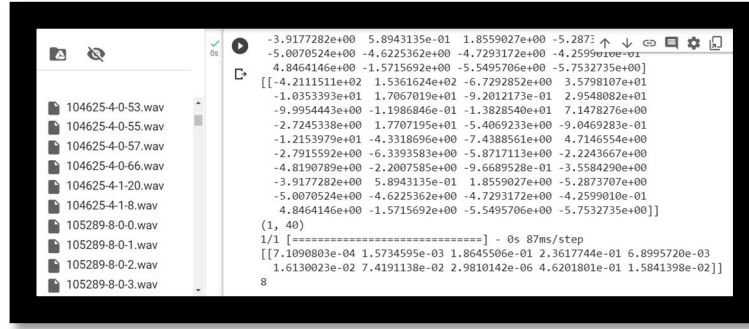


Figure 7. Testing the model

TABLE II. CSV FILE OF DATASET – SIREN

slice_file_name	fsID	start	end	salience	fold	classID	class
102853-8-0-0.wav	102853	0	4	2	7	8	siren
102853-8-0-1.wav	102853	0.5	4.5	2	7	8	siren
102853-8-0-2.wav	102853	1	5	2	7	8	siren
102853-8-0-3.wav	102853	1.5	5.5	2	7	8	siren
102853-8-0-4.wav	102853	2	6	2	7	8	siren
102853-8-0-5.wav	102853	2.5	6.5	2	7	8	siren
105289-8-0-0.wav	105289	26.31554	30.31554	2	7	8	siren
105289-8-0-1.wav	105289	26.81554	30.81554	2	7	8	siren
105289-8-0-2.wav	105289	27.31554	31.31554	2	7	8	siren
105289-8-0-3.wav	105289	27.81554	31.81554	2	7	8	siren
105289-8-0-4.wav	105289	28.31554	32.31554	2	7	8	siren
105289-8-0-5.wav	105289	28.81554	32.81554	2	7	8	siren
102871-8-0-12.wav	102871	6	10	1	2	8	siren
102871-8-0-13.wav	102871	6.5	10.5	1	2	8	siren
102871-8-0-14.wav	102871	7	11	1	2	8	siren
102871-8-0-15.wav	102871	7.5	11.5	1	2	8	siren
102871-8-0-2.wav	102871	1	5	1	2	8	siren
102871-8-0-3.wav	102871	1.5	5.5	1	2	8	siren
102871-8-0-4.wav	102871	2	6	1	2	8	siren
102871-8-0-5.wav	102871	2.5	6.5	1	2	8	siren
102871-8-0-6.wav	102871	3	7	1	2	8	siren
102871-8-0-7.wav	102871	3.5	7.5	1	2	8	siren

VI. COMPARATIVE STUDY

TABLE III. COMPARATIVE STUDY OF REFERENCE PAPERS AND PROPOSED METHOD

Study	Features	Method	Accuracy
Michela Cantarini , Leonardo Gabrielli and Stefano Squartini (2022) [1]	Few-Shot Metric Learning Feature	CNN	87% - 91%
Gour Karmakar, Abdullahi Chowdhury, Joarder Kamruzzaman, Iqbal Gondal (2021) [2]	Severity of an accident	Priority based code and clearance code	70%

Rubens Cruz Gatto and Carlos Henrique Quartucci Forster (2021) [3]	MFCC, Video recording	OpenCV and ML model	74%
Proposed ML model	MFCC	ANN	85% - 90%

Table III represents that the proposed model is better in accuracy and extracting features compared to other reference papers. In [1], the feature extracted is different than MFCC which gave them better accuracy with the CNN model proposed. But it has been concluded that few-shot metric and MFCC can be very crucial for audio feature extraction and pre-processing techniques compared to other random sampling data points.

VII. CONCLUSION

The ML model which was trained to classify the audio data into 'siren' or not has an accuracy of 90%. The model successfully classified the siren audio data as a class of siren and as the model is trained with some noise data of traffic and public noise the model classifies as them to their respective classes too. The MFCC feature extracted provided with actual accurate data points which helped the model to map with their class. And it can be concluded that a high-grade feature extraction method can be combined with simple ANN architecture and produce accurate prediction results to classify audio data or siren sounds. The novelty of this project is that it is unlike any other methods involving image processing in CCTV cameras in traffic signals which is very much inefficient to identify the type of the vehicle with simple networks. And knowing the arrival of ambulance when it is near the traffic signal is unlikely for the traffic to make emergency measures to move and control the traffic signal lights. And it is much reliable when it comes to audio recognition of ambulance before-hand than antenna (RFID) or image processing-based system. The scope of this project is it could widely contribute to save human lives in which even minutes matter and can help the traffic system with an emergency-alert based system which could be implemented for various other emergency or V.I.P vehicles. So, the future work includes creating an Ambulance Alert System with the help of IoT and embedded systems which can be implemented in traffic junctions that acts upon detection of ambulance siren and alert the junction of the arrival and managing the traffic lights during that emergency.

REFERENCES

- [1] Michela Cantarini, Leonardo Gabrielli and Stefano Squartini "Few-Shot Emergency Siren Detection" - Sensors Journal 2022, VOL 22 4338, <https://doi.org/10.3390/s22124338>
- [2] Gour Karmakar, Abdullahi Chowdhury, Joarder Kamruzzaman, Iqbal Gondal, "A Smart Priority-Based Traffic Control System for Emergency Vehicles", IEEE SENSORS JOURNAL, Volume: 21, Issue: 14 JULY 15, 2021
- [3] Rubens Cruz Gatto and Carlos Henrique Quartucci Forster, "Audio-Based Machine Learning Model for Traffic Congestion Detection", IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 22, NO. 11, NOVEMBER 2021
- [4] Van-Thuan Tran and Wei-Ho Tsai, "Audio-Vision Emergency Vehicle Detection", IEEE SENSORS JOURNAL, DECEMBER 15, 2021, VOL. 21, NO. 24
- [5] Mohamed Abdel-Aty, Zubayer Islam, "Real-time Emergency Vehicle Event Detection Using Audio Data", IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, 2022
- [6] Van-Linh Nguyen, Ren-Hung Hwang and Po-Ching Lin, "Controllable Path Planning and Traffic Scheduling for Emergency Services in the Internet of Vehicles", IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, AUGUST 2022
- [7] Dataset : <https://urbansounddataset.weebly.com/urbansound8k.html>
- [8] Jannik Zörn and Wolfram Burgard, "Self-Supervised Moving Vehicle Detection From Audio-Visual Cues", IEEE ROBOTICS AND AUTOMATION LETTERS, VOL. 7, NO. 3, JULY 2022.
- [9] Li Z, Shahidehpour M, Bahramirad S & Khodaei A, "Optimizing traffic signal settings in smart cities", IEEE Transactions on Smart Grid, Vol.8, No.5, (2017), pp.2382-2393.
- [10] Dharma Rane, Pushkar Shirodkar, Trilochan Panigrahi, and S. Mini, "Detection of Ambulance Siren in Traffic", International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), MARCH 2019 INSPEC Accession Number: 19453701