International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

Ambulance Siren Detection using Artificial Neural Network

Deepak Kadam[1], Nisarg Patel[2], Meher Patil[3], Prof. G. B. Aochar[4]

[1-4]*Department of Computer Engineering, Modern Education Society's College of Engineering, Pune, Maharashtra, India*

Abstract - *Due to ever-increasing population and people's rising disposable income, vehicles are becoming omnipresent, especially in major cities around the world. Most cities were not developed considering future vehicular population. This has led to increasingly jammed roads all over the cities, and poor traffic management may lead to catastrophic consequences. Emergency vehicles, like ambulances and fire trucks, are the most affected ones due to traffic problems, as lives could be at stake. We aim to solve this problem using Audio fingerprinting which is a technique to identify types sound patters by comparing them with available samples of similar patterns, similar to how traditional fingerprint identification works. This will be done by manipulating the traffic signals based on audio evidence of a waiting emergency vehicle which will be detected using its distinctive siren sound patterns. This distinctiveness of the siren sounds will be used for the audio fingerprinting purpose.*

Key Words: **Ambulance siren, Emergency vehicle, Audio fingerprinting, Traffic management, Machine Learning, Deep Learning, Artificial neural network, Logistic Regression.**

1. INTRODUCTION

Emergency vehicles on active duty often get stuck at traffic signals. If the number of vehicles at the signal is large, the emergency vehicles may have to wait for more than one signal cycle. Usually, there is little space for the other vehicles to move to make way for the emergency vehicles. All these situations may have serious consequences if a fire truck or an ambulance fails to reach its destination in time. A good traffic management system is required to prevent such issues and our system is designed to complement the traffic management system.

We studied a few different methods that have been proposed for the purpose of helping emergency vehicles avoid traffic congestions, without the use of sound. These methods have limitations that we intend to solve with our system. In [1] a Radio Frequency Identification (RFID) based system is proposed in which the density of vehicles is determined to decide whether or not to change the traffic signal to green. This method also proposes use of RFIDs of different ranges in order to detect emergency vehicles and clear the traffic on their detection. This method has significant limitations. Firstly, if the density of vehicles is not adequate then the signal will not turn green and the emergency vehicles will remain stuck at the signal. Secondly, if each emergency vehicle is fitted with a unique RFID tag, it will need to be very close to the RFID scanner in order to get detected, which is not always possible if there is not enough space for the vehicle to maneuver.

In [2] an image processing method has been proposed that uses convolutional neural networks to detect emergency vehicles in images of traffic sourced from CCTV cameras. The main limitation of this method is that it will fail to detect emergency vehicles that are behind larger vehicles such as a bus or a trailer truck. Other significant limitation is the relatively high cost of the equipment needed for this system.

We studied a few proposed sound-based emergency detection methods as well. But these also have certain limitations that we intend to solve with our system. In [3] detection of siren sound is done using a modified minimum mean square error method which uses the peaks of the sound wave frequency. This method uses the distinct frequencies of siren sounds as its base for detection. The limitation of this method is that there may exist sounds having frequencies similar to those of the siren sounds of emergency vehicles. In such cases, the system will falsely identify a non-siren sound as siren. Our method overcomes this limitation by using a large number of real audio samples of siren sounds with which an input sample is matched to detect sound patterns similar to those present in the real audio samples.

In [4] a method similar to ours is proposed but it uses only clear siren sound samples and not real-time traffic noise. Our method, on the other hand, uses a pre-processing mechanism to extract siren sounds, if any, from raw traffic sound recordings and reduce noise from it as much as possible. This resulted in significant improvement in the accuracy of siren detection by our model, compared to the other method.

We have used a combination of pre-processing and Artificial Neural Network (ANN) based logistic regression to extract and detect siren sounds of emergency vehicles from real-time traffic noise. The traffic noise will be recorded by an array of sensitive microphones and the audio will be divided into clips of fixed short durations for processing. We have limited our testing to only ambulance sirens but our method can be applied to detect fire trucks and other emergency vehicles as well, without any structural modifications to the algorithm.

2. METHODOLOGY

2.1 Pre-processing

Raw traffic noise audio from the microphone array is provided to the pre-processing module as input. This module uses a sample of several different clear siren sounds as a

second input. Both inputs are converted to frequency-domain, from time-domain, using fast Fourier Transform. Noise is obtained by subtracting the common numerical data among the two inputs, if any, from the raw traffic audio. We call this process as masking. The obtained noise consists of all the traffic sounds except any siren sound which was removed in the earlier step. This noise is then subtracted from the original raw traffic audio. This will result in extracted siren sounds, if any, from the raw traffic audio with reduced noise. This output is then provided as input to the detection module.

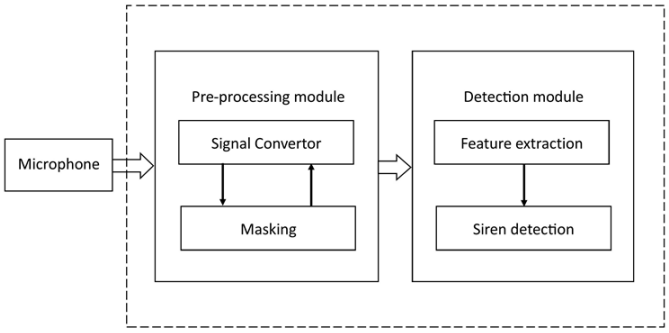

Fig-1: Architecture of our siren detection system

2.2 Detection

2.2.1 Studying the audio

Before the detection process, study of audio signal is very important. The study involves analysis of signal, extracting its properties, predicting its behaviour, finding out if any specific pattern is present in the signal, etc. In our system, the input audio signal has environmental sounds like vehicle horns, engine sounds, human noise, etc. along with siren sounds of emergency vehicles, if any. Over the years, audio signal processing has grown significantly in the way of signal analysis and classification. Audio signals cannot be used directly for a machine learning (ML) algorithm. This issue can be solved by integrating the modern ML techniques with feature extraction methods.

2.2.2 Feature extraction

The performance metrics of any ML algorithm depends on how accurately the expected features are extracted from the audio signals. This method provides precious data from signal in an understandable format for machine learning models. Feature extraction is required for classification, prediction, and recommendation algorithms.

We studied three ways of feature extraction, namely zero crossings, spectral centroid and Mel-Frequency Cepstral Coefficient (MFCC). For our system, we have used MFCC based feature extraction as it is one of the most important and standard methods to extract features from audio signals and is majorly used whenever working with audio signals, especially in ML applications.[5] The MFCCs of a signal are a small set of features (usually about 10–20) which concisely describe the overall shape of a spectral envelope. They are based on the nonlinear frequency feature of human ears. In essence, MFCC works by selecting energy in different frequency bands as the feature of target.

2.2.3 Handling of dataset

Categorical data is the data that generally takes a limited number of possible values. All machine learning models are some kind of mathematical model that need numbers to work with. This is one of the primary reasons we need to pre-process the categorical data before we feed it to an ML model. The data used in our system is in the form of audio files but a machine cannot understand such data. It has to be transformed into machine-understandable format like numerical, binary, etc. So, emergency siren sound is represented as 1 and non-emergency sounds as 0.

2.2.4 Technique used for detection of siren

We have implemented a Multi-Layer Perceptron (MLP) model for the detection purpose. The model builds upon thousands of training samples consisting of siren as well non-siren audio. MLP is essentially a logistic regressor, implemented using a feedforward ANN, which uses features, extracted audio features in our case, to determine whether the input audio contains siren or not, based on learned information.

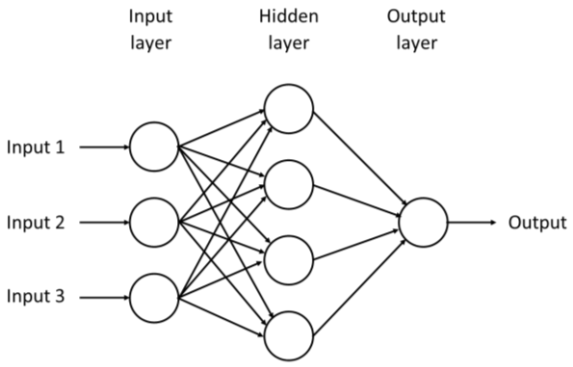

Fig-2: Multi-Layer Perceptron

In the MLP, there can be more than one linear layer i.e. combinations of neurons. It has three or more layers. It is used to classify data that cannot be separated linearly. It is a type of fully connected ANN in which every single node in a layer is connected to each node in the following layer. An MLP uses a nonlinear activation. We have used the Rectified Linear Unit (ReLU) activation function for the hidden layer and Sigmoid Function for the output layer. ReLU applies a non-saturating activation function. It effectively removes negative values from an activation map by setting them to zero. It increases the nonlinear properties of the decision function and of the overall network without affecting the receptive fields of the convolution layer. Other functions are also used to increase nonlinearity, for example the saturating hyperbolic tangent and the sigmoid function. ReLU is often preferred to other functions because it trains the neural

network several times faster without a significant penalty to generalization accuracy.

Our model uses the Root Mean Square Propagation (RMSprop) optimizer for the learning process. It restricts the oscillations in the vertical direction. Therefore, we can increase our learning rate and our algorithm can take larger steps in the horizontal direction, converging faster. Binary Cross Entropy loss function is used to estimate the loss of the model so that the weights can be updated to reduce the loss on the next evaluation. It belongs to probabilistic class. It computes the cross-entropy loss between true labels and predicted labels. Used when there are only two label classes (assumed to be 0 and 1).

3. RESULTS

For samples of real-world traffic sounds, our system was able to achieve high accuracy of detecting both presence and absence of siren. We also observed that accuracy of detection was significantly less for samples of raw traffic audio than for pre-processed audio samples. The final observed outcomes of our system are tabulated below.

Table-1: Results

Input	Expected output	Actual output
Traffic audio, raw as well as pre-processed, not containing siren sound	Siren not detected. High accuracy	Siren not detected. Accuracy ≥ 98%
Pre-processed traffic audio containing siren sound	Siren detected. High accuracy	Siren detected. Accuracy ≥ 95%
Raw traffic audio recording containing siren sound	Low accuracy	Accuracy ≤ 85%

The Confusion Matrix is given below.
EM – Emergency
Non-EM – Non-Emergency

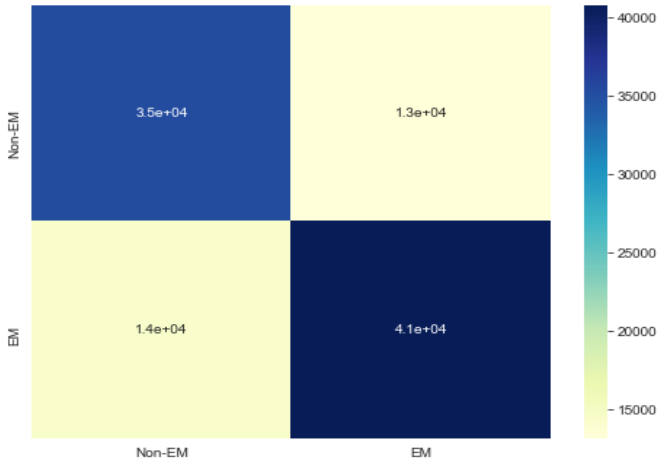

Fig-3: Confusion Matrix

4. CONCLUSION AND FUTURE WORK

In conclusion, we propose this method for detection of emergency vehicles stuck at traffic signals as it overcomes the limitations of the other methods as well as for its simplicity and cost effectiveness. The scope of our system is currently limited only to the detection of siren. As future work, we intend to extend our system to include tracking of the detected emergency vehicle to ensure its passage through the traffic junction in order to revert the traffic signal to its default behavior. We also intend to further improve the accuracy of our detection system by implementing a better algorithm.

REFERENCES

[1] T. Naik, Roopalakshmi R., Divya Ravi N., P. Jain, Sowmya B. H., Manichandra. "RFID-Based Smart Traffic Control Framework for Emergency Vehicles". Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018.

[2] S. Roy, Md. Sakif Rahman. "Emergency Vehicle Detection on Heavy Traffic Road from CCTV Footage Using Deep Convolutional Neural Network". International Conference on Electrical, Computer and Communication Engineering (ECCE), 2019.

[3] Kiran S. L., Supriya M. "Siren Detection and Driver Assistance using Modified Minimum Mean Square Error Method". International Conference on Smart Technology for Smart Nation, 2017.

[4] Van-Thuan Tran, Yu-Cheng Yan, Wei-Ho Tsai. "Detection of Ambulance and Fire Truck Siren Sounds using Neural Networks". Proceedings of 51st Research World International Conference, Hanoi, Vietnam, 2018.

[5] P. P. Singh, P. Rani. "An Approach to Extract Feature using MFCC". IOSRJEN, vol 4, no 8, pp 21, 2014.