RESSTECH

Vol. 1, No. 1, pp. 01-08, 2024

RESEARCH ON INSTRUMENTATION

https://rin.resstech.org/

Comparative Analysis of Siren Classification Technique for Emergency Vehicles

Ei Paing Phyo*, Hla Myo Tun, Thanda Win and Lei Lei Yin Win

Department of Electronic Engineering, Yangon Technological University, Yangon, MYANMAR

**Corresponding author: eipaiphyo@gmail.com*

ARTICLE INFO

Article history:
Received Mei 22, 2024
Revised Mei 29, 2024
Accepted June 14, 2024
Available online June 30, 2024

Keywords*:*
Audio Signal Processing
Hybrid Method
Machine Learning Classifier
Siren Classification

ABSTRACT

Emergency vehicle sirens greatly aid traffic control and public safety awareness. Improving emergency response systems requires accurate siren classification. This study aims to categorize emergency vehicles, particularly fire trucks, police cars, and ambulances, based on the features of their sirens. It thoroughly analyses various schemes for categorizing emergency vehicle sirens. Mel-Frequency Cepstral Coefficients (MFCC), Zero-Crossing Rate (ZCR), Spectral Centroid, and hybrid methods that combine MFCC with ZCR and Spectral Centroid were observed for comparison. The data set is sourced from the Google Audio Set Ontology, ensuring robust training and evaluation of the models. This methodology involves preprocessing audio data, extracting relevant features, and training classifiers. The proposed hybrid method combines MFCC with Spectral Centroid to leverage their complementary strengths. Through rigorous experimentation, this system evaluates the performance of different classifiers, aiming to provide insights for optimal siren classification. The findings contribute to advancing audio classification methodologies and have implications for developing more robust emergency response and traffic management systems.

1. INTRODUCTION

Emergency vehicle sirens produce distinct sound signals essential for efficient traffic control and public safety awareness. Accurate recognition and categorization of these signals are crucial for enhancing the emergency response system. This study thoroughly examines various approaches to classify emergency vehicle sirens. It seeks to evaluate and compare the performance of different feature extraction techniques. In addition, this study aims to improve the robustness of real-time classification by hybridizing feature extraction methods. Multiple features in sound signals can help in analysis. The CNN solution may achieve high accuracy rates and lower error values. However, implementing it requires larger feature datasets and more extensive data processing [1]. The focus of the emergency vehicle recognition system is not on cost-friendly but on saving the lives of people. So, sound and image recognition techniques [2] control tooling functions, although the sensors above the emergency vehicles must be fixed [3].

Furthermore, while the model has become more accurate and street noise has decreased, it remains crucial to consider the safety concerns of individuals with hearing impairments [4]. Acoustic detection depends on the amount of noise. In the research field, weak detected signals can be affected by heavy road noise or other artificial noise [5]. To improve accuracy, Signal and image processing utilize various noise reduction techniques and hybrid noise removal methods [6].

This study explicitly contrasts the Zero-Crossing Rate (ZCR), Spectral Centroid, and Mel-Frequency Cepstral Coefficients (MFCC) and hybrid techniques that integrate MFCC with ZCR and Spectral Centroid. Given the primary objective of real-time implementation, the system necessitates simplicity and efficiency. Various classification algorithms of K-Nearest Neighbors (KNN) classifiers, Decision Tree classifiers, Support Vector Machine (SVM) classifiers, and Naive Bayes classifiers can be utilized to evaluate the accuracy [7].

2. MATERIAL AND METHOD

The system evaluates the performance of several feature extraction techniques and classifiers using systematic technologies. The dataset, a sizable tagging collection of labelled audio samples covering many sound categories, including emergency vehicle sirens, is sourced from the Google Audio Set Ontology[8]. This data set contains videos of the ambulance siren of 1939, the police car siren of 3659, the fire engine, and the fire truck siren of 3199[9]. By utilizing this dataset, this system can thoroughly train and assess classification models, guaranteeing their resilience and applicability. First, the audio data undergoes preprocessing, and then feature extraction techniques such as MFCC, ZCR, Spectral Centroid, and hybrid methods are utilized to extract pertinent features. In MFCC, the audio data undergoes loading and normalization to ensure its values are in the range of [-1,1]. Subsequently, resample the signal to a standard 44.1kHz sampling rate. Then, the system extracts 13 mel-frequency cepstral coefficients from the resampled signal [10]. The MFCC technique initially converts audio from analogue format region to digital and applies pre-emphasis using a first-order high pass filter. Subsequently, a hamming window is applied to segment the signal, effectively reducing noise in the high-frequency range. Analyzing the audio signal is faster in the frequency domain than the time domain, so Direct Fourier Transform is employed to convert the signal to the frequency domain. A mel-filter bank is needed to pass the signal to improve the performance and to map the mel-scale. Then, the output is applied to log and inverse transform as a preprocessing stage[11]. In the preprocessing case of ZCR, the audio file is read and converted from stereo to mono. Establishing the zero-crossing rate for each signal frame involves using a window length of 30 milliseconds and a hop size of 15 milliseconds. Subsequently, it calculates the collective ZCR by averaging all individual frame-wise ZCR values. The spectral centroid feature extraction method requires loading and normalizing the audio file. This process is similar to MFCC, which aids in maintaining consistent processing, mainly when dealing with audio files of varying amplitudes. The function needs to calculate the spectral centroid feature for the audio signal. For consistency, all feature extraction methods use audio segmentation of 3s duration.

According to the sampling theorem, a signal must have a minimum sample rate twice its highest frequency to adequately represent it without losing information. Most people agree that the human hearing range extends up to 20kHz. Thus, for audio signals up to 22.05kHz in frequency, a sampling rate of 44.1kHz meets the Nyquist requirement. Moreover, it is CD standard and works with various software and audio devices[12].

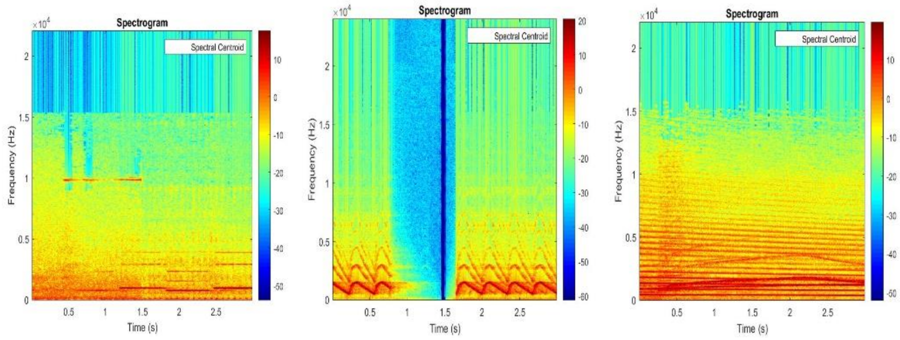
This research used MATLAB R3019a as a software tool [13] and four kinds of classifiers. KNN is a non-parametric classifier that determines the label for unknown data points by utilizing the majority class of its k nearest neighbours. In this study, the value of k is 3. The SVM-supervised learning algorithm finds the hyperplane that best divides classes in high-dimensional feature space. It seeks to reduce classification error while maximizing the margin between classes. The decision tree derives its conclusions from the values of the features at the internal nodes, segmenting the feature spaces recursively into regions. The Bayes theorem-based Naive Bayes classifier is a straightforward yet effective probabilistic algorithm that relies on the premise of feature independence.

A frequency-modulated waveform, in which the frequency varies periodically, is a characteristic of a typical siren [14]. The main spectral component in a siren signal is the one that corresponds to the lowest frequency, even though it has multiple harmonic components [15]. Thus, this system describes the general form of the spectral envelope in the frequency domain using MFCC [16]. ZCR measures the rate at which a signal transitions between positive, negative, and zero values [17] and spectral centroid to characterize the center of mass of a sound's spectrum [18]. Most audio classification systems employ the MFCC technique with a support vector machine (SVM) in binary classification form [19]. In addition, the system checks the accuracy of the classifiers, which can be affected by variable input duration [20]. The efficacy of these classifiers in audio classification tasks is the reason behind their selection.

Table 1 Data set of siren sound signal

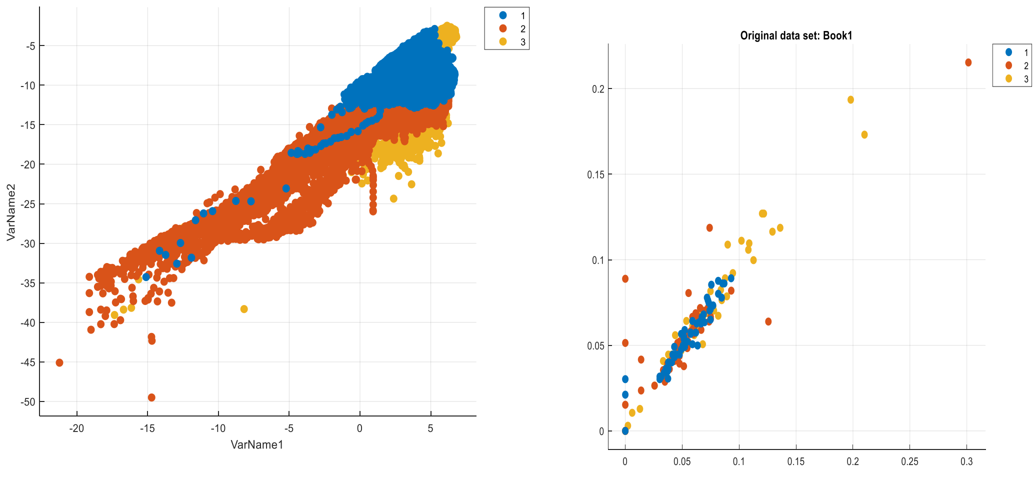
Data Set	Google Audio Set Ontology
Classes	Ambulance, Police Car, Fire Truck
Average Clip Duration	3 seconds
Sampling Rate	44.1kHz

Figure 1 displays the spectrograms of three distinct sirens with classification based on spectrogram analysis [21] [22].

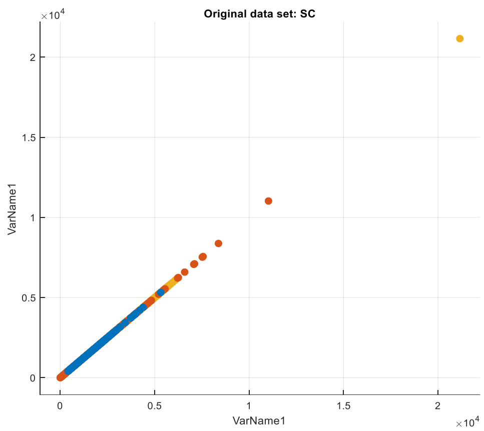


(a) (b) (c)

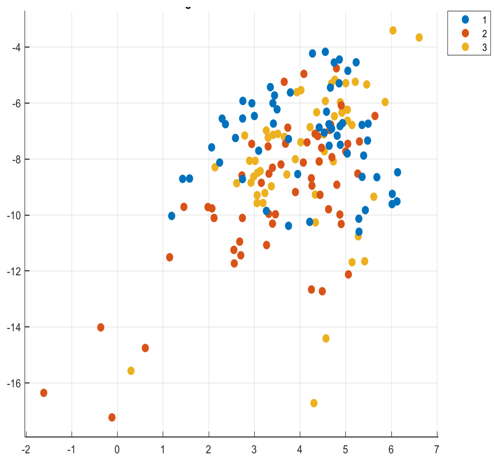
Figure 1. Spectrograms of three types of siren sounds; (a) ambulance siren, (b) police car siren, and (c) fire truck siren



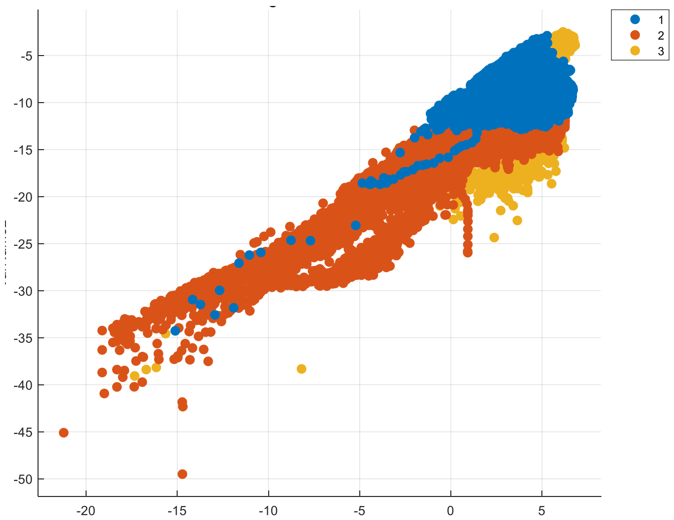
(a) (b)



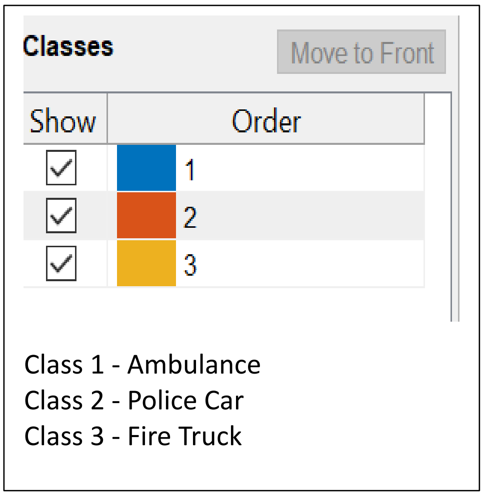
(c)



(d)



(e)



(f)

Figure 2. Original data sets, features extracted by Figure (a)MFCC, Figure (b) ZCR, Figure (c) Spectral Centroid, Figure (d) Hybrid method of MFCC and ZCR, Figure (e) Hybrid method of MFCC and Spectral Centroid and Figure (f) Classes

3. RESULTS AND DISCUSSION

$$ZCR = \frac{1}{N} \sum_{n=1}^{N-1} |sgn(Signal(n) - sgn\big(Signal(n-1)\big)| \quad (1)$$

$$SC = \frac{\sum_{k=0}^{N-1} f(k).P(k)}{\sum_{k=0}^{N-1} P(k)} \quad (2)$$

ZCR is zero crossing rate, and N is the total number of detected samples in the audio signal. Signal (n) represents the audio signal amplitude at sample n, and the signum function, sgn (x), returns -1 in the case of a negative x, 0 in the case of a zero x, and 1 in the case of a positive x. S.C. stands for spectral centroid, f(k) is the frequency of the kth bin, P(k) is the magnitude of the kth bin, and N is the total number of bins in the spectrum.

Bar charts depicted this research's varied results. As seen in Figures 4 and 5, these charts represent the various feature extraction techniques and classification accuracy attained by testing them using four classifiers. Additionally, training times for the four classifiers were examined.

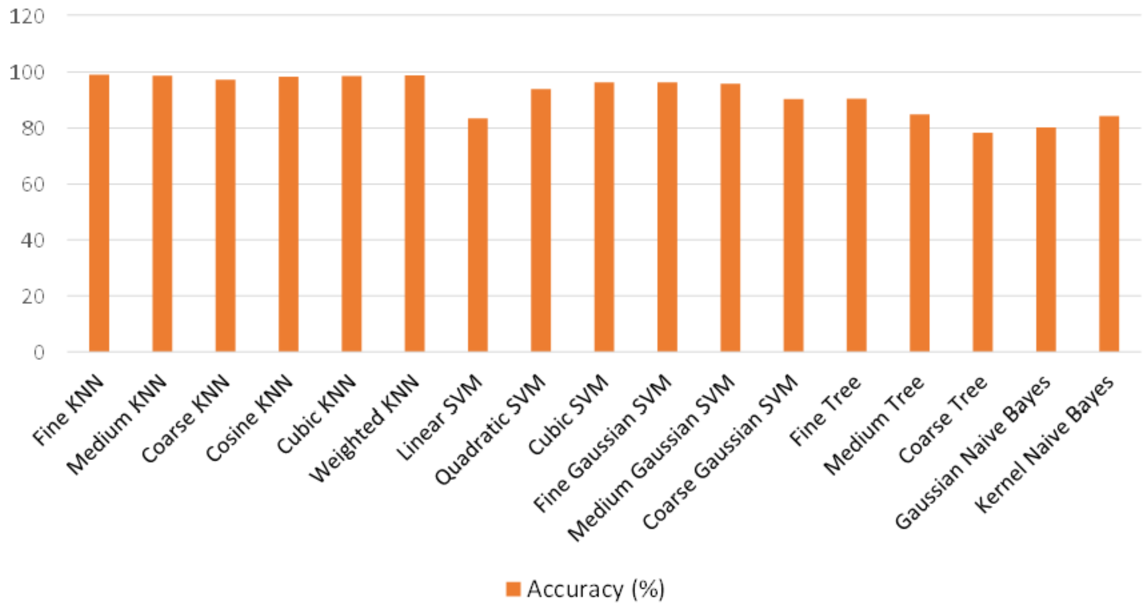


Figure 3. Classification accuracy of four classifiers using MFCC feature extraction

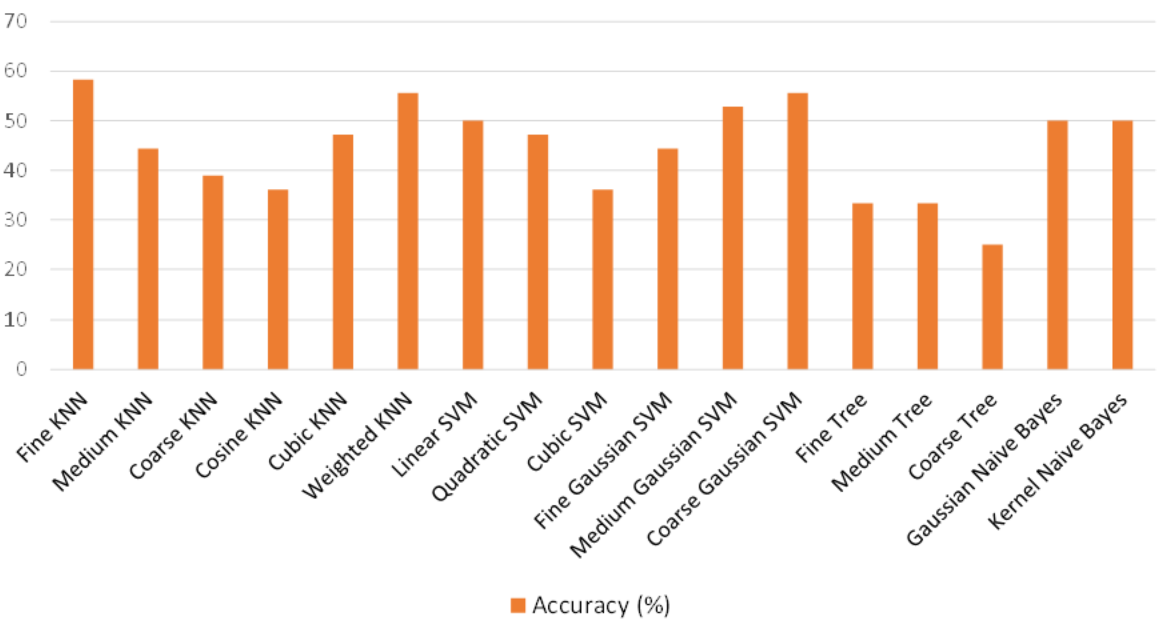


Figure 4. Classification accuracy of four classifiers using ZCR feature extraction

4. CONCLUSION

In conclusion, this study indicates that the hybridization of MFCC and spectral methods, coupled with the KNN classifier, represents the most suitable approach for audio classification tasks. The KNN classifier employs the majority voting mechanism of the k nearest neighbors and serves as the primary tool for classifying different types of sound input. These results provide crucial new information for creating precise and effective categorization systems. These results provide valuable insights for practical improvements in audio analysis and for enhancing classification systems. These findings seek to improve public safety and emergency response, optimize resource use, and be integrated with other systems to offer an all-encompassing emergency management solution[26]. The KNN classifier emerges as a top choice for real-time classification, consistently surpassing alternative methods. Thus, this research idea holds promise for guiding the development of efficient siren classification systems capable of swift data processing. The research findings have implications beyond siren analysis with potential application in fields such as speech recognition, music processing, and environmental sound classification. Most of the research tried to reduce noise at the preprocessing stage to get high accuracy. This study proves that hybridization enhances accuracy and reduces computational complexity. This point is one of the research's highlights. Furthermore, the study aids in optimizing system resources by identifying highly accurate methods, thereby making it more feasible to deploy siren classification systems in resource-constrained environments. Overall, this research serves as a valuable foundation for further studies in audio classification and offers practical insight for researchers, practitioners, and developers seeking to implement effective and efficient siren analysis solutions.

ACKNOWLEDGMENTS

The author extends special thanks to all colleagues from the Signal Processing Research Group under the Department of Electronic Engineering at Yangon Technological University for their contributions to theoretical methods and sincerely thanks everyone who offered direct or indirect support in completing this research.

DECLARATIONS

Authorship contribution

Ei Paing Phyo: conceiving and designing the research study, collecting and analysing data, and writing the article. **Thanda Win**: supervision. **Lei Lei Yin Win**: providing critical revisions to the article. **Hla Myo Tun**: reviewing and editing.

Competing Interest

The authors declare that they have no competing interests that could influence the conduct or reporting of this research work.

Funding statement

No particular public, private, or nonprofit grant supported this research.

Ethical Clearance

There are no human subjects in this manuscript, and informed consent is not applicable.

REFERENCES

[1] K. Kumar and K. Chaturvedi, “An Audio Classification Approach using Feature Extraction Neural Network Classification Approach,” 2020.

[2] H. Razalli, R. Ramli, and M. H. Alkawaz, “Emergency Vehicle Recognition and Classification Method Using HSV Color Segmentation,” in Proceedings - 16th IEEE International Colloquium on Signal Processing and its Applications Feb. 2020, pp. 284–289. doi: 10.1109/CSPA48992.2020.9068695.

[3] T. Swathi and V. Mallu, “Emergency Vehicle Recognition System,” International Journal of Engineering Trends and Technology (IJETT), 2013, [Online]. Available: http://www.ijettjournal.org

[4] A. Esquivel Ramirez and E. Donati, “A siren identification system using deep learning to aid hearing-impaired people.”

[5] T. Chen, D. Yu, B. Wu, and B. Xia, “Weak Signals Detection by Acoustic Metamaterials-Based Sensor,” IEEE Sens J, vol. 21, no. 15, pp. 16815–16825, Aug. 2021, doi: 10.1109/JSEN.2021.3076860.

[6] H. Alaeiyan, M. R. Mosavi, and A. Ayatollahi, “Hybrid noise removal to improve the accuracy of inertial sensors using lifting wavelet transform optimized by genetic algorithm,” Alexandria Engineering Journal, vol. 80, pp. 326–341, Oct. 2023, doi: 10.1016/j.aej.2023.08.069.

[7] S. L. Ullo, S. K. Khare, V. Bajaj, and G. R. Sinha, “Hybrid Computerized Method for Environmental Sound Classification,” IEEE Access, vol. 8, pp. 124055–124065, 2020, doi: 10.1109/ACCESS.2020.3006082.

[8] U. Mittal and P. Chawla, “Acoustic Based Emergency Vehicle Detection Using Ensemble of Deep Learning Models,” in Procedia Computer Science, Elsevier B.V., 2022, pp. 227–234. doi: 10.1016/j.procs.2023.01.005.

[9] “Google Audio Set Ontology”, Accessed: May 10, 2024. [Online]. Available: https://research.google.com/audioset/dataset/index.html

[10] A. J. Abougarair, M. K. Aburakhis, and M. O. Zaroug, “Design and implementation of smart voice assistant and recognizing academic words,” International Robotics & Automation Journal, vol. 8, no. 1, pp. 27–32, Feb. 2022, doi: 10.15406/iratj.2022.08.00240.

[11] M. A. Suhaimy, I. S. A. Halim, S. L. M. Hassan, and A. Saparon, “Classification of ambulance siren sound with MFCC-SVM,” in AIP Conference Proceedings, American Institute of Physics Inc., Dec. 2020. doi: 10.1063/5.0032392.

[12] R. Arie, A. Brand, and S. Engelberg, “IEEE Instrumentation & Measurement Magazine Compressive Sensing and Sub-Nyquist Sampling,” 2020.

[13] Y. Sanghani, “Audio Signal Processing using Matlab”, doi: 10.13140/RG.2.2.19169.79202.

[14] M. Cantarini, A. Brocanelli, L. Gabrielli, and S. Squartini, “Acoustic features for deep learning-based models for emergency siren detection: An evaluation study,” in International Symposium on Image and Signal Processing and Analysis, 2021. doi: 10.1109/ISPA52656.2021.9552140.

[15] D. K. Fragoulis and J. N. Avaritsiotis, “A siren detection system based on mechanical resonant filters,” Sensors, vol. 1, no. 4, pp. 121–137, 2001, doi: 10.3390/s10400121.

[16] Z. K. Abdul and A. K. Al-Talabani, “Mel Frequency Cepstral Coefficient and its Applications: A Review,” IEEE Access, vol. 10, pp. 122136–122158, 2022. doi: 10.1109/ACCESS.2022.3223444.

[17] IEEE Robotics and Automation Society and Institute of Electrical and Electronics Engineers, ICRA2017 : IEEE International Conference on Robotics and Automation : program : May 29-June 3, 2017, Singapore.

[18] T.-T. Lie, Y. Liu, Hong Kong Society of Mechanical Engineers, IEEE Power & Energy Society, and Institute of Electrical and Electronics Engineers, 2020 5th Asia Conference on Power and Electrical Engineering (ACPEE 2020) : proceedings : 4-7 June 2020, Chengdu, China.

[19] S. M. Kamruzzaman, A. N. M. R. Karim, and M. E. Haque, “Speaker Identification using MFCC-Domain Support Vector Machine.”

[20] V. T. Tran and W. H. Tsai, “Acoustic-Based Emergency Vehicle Detection Using Convolutional Neural Networks,” IEEE Access, vol. 8, pp. 75702–75713, 2020, doi: 10.1109/ACCESS.2020.2988986.

[21] D. Pramanick, H. Ansar, H. Kumar, S. Pranav, R. Tengshe, and B. Fatimah, “Deep learning based urban sound classification and ambulance siren detector using spectrogram,” in 2021 12th International Conference on Computing Communication and Networking Technologies, INT 2021, 2021. doi: 10.1109/ICCCNT51525.2021.9579778.
[22] L. Marchegiani and P. Newman, “Listening for Sirens: Locating and Classifying Acoustic Alarms in City Scenes,” IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 10, 2022, doi: 10.1109/TITS.2022.3158076.
[23] IEEE Signal Processing Society, 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing: proceedings: May 26-31, 2013, Vancouver Convention Center, Vancouver, British Columbia, Canada.
[24] Amity University, Institute of Electrical and Electronics Engineers, and Institute of Electrical and Electronics Engineers. United Kingdom and Republic of Ireland Section, 2020 International Conference on Intelligent Engineering and Management (ICIEM-2020).
[25] Dayananda Sagar College of Engineering, Institute of Electrical and Electronics Engineers. Bangalore Section, and Institute of Electrical and Electronics Engineers, 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2020): conference proceeding: 5-7 March 2020.
[26] A. Manoj Kumar, G. Sai Chandan Reddy, G. Hari Krishna, and E. Bhanu Venkata Avinash, “Emergency Vehicle Detection in Heavy Traffic Using Computer Vision,” 2023. [Online]. Available: www.ijrpr.com