IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Real-Time Victim Audio Detection System For Reducing Crime Rate Using Ml And Dl

[1]Mrs. Sk. Sameerunnisa, [2]Chinthala Akhil, [3]Kandru Prabhushan, [4]Kosuri Vedhavathi,[5]Motupalli Jyothi

[1]Assistant Professor, [2]Student, [3]Student, [4]Student, [5]Student,

[1,2,3,4,5]Department of CSE (IoT, Cybersecurity including Blockchain Technology),

[1,2,3,4,5]Vasireddy Venkatadri Institute of Technology, Nambur, India.

Abstract—The escalating frequency of crimes involving victims highlights the urgent demand for technological advancements to enhance public safety. This research presents a Real-Time Victim Audio Detection System based on Machine Learning (ML) and Deep Learning (DL) techniques. The innovative technology is designed to recognize distress calls from audio inputs, enabling immediate alerts to law enforcement for swift response. The system integrates audio feature extraction techniques, such as Mel-Frequency Cepstral Coefficients (MFCC) and spectrogram analysis, to classify audio signals and identify distress situations. The deep learning model, trained on a comprehensive data set including genuine and simulated distress audio samples, demonstrates high accuracy and robustness across various environments. The system's real-time capabilities ensure rapid detection and reporting, reducing response times and potentially preventing criminal acts. This innovative approach exemplifies the application of artificial intelligence to improve public safety by addressing the limitations of traditional surveillance and monitoring systems.

Index Terms— Audio Detection, Crime Prevention, Real-Time Systems, Machine Learning, Deep Learning, Distress Signal Recognition, Automated Alerts.

I. INTRODUCTION

The global rise in criminal activities has made public safety a top priority for governments, law enforcement, and communities. To combat this escalating issue, the creation and deployment of efficient surveillance systems have become crucial. While conventional surveillance methods like CCTV cameras and manual monitoring have been instrumental in detecting criminal behavior, they have inherent shortcomings. These include delayed reactions, limited coverage, and a lack of proactive capabilities[7]. For example, CCTV cameras are restricted to capturing visual data within their specific range, creating blind spots in certain areas, and depend on human operators who may overlook critical incidents[13].

Furthermore, relying on human assessment to review extensive footage can lead to slow response times, which can be vital in crime prevention or mitigation[1].To tackle these issues, it is necessary to incorporate more sophisticated technologies that can surmount these limitations by enabling instantaneous analysis and action. In the last ten years, progress in Machine Learning (ML) and Deep Learning (DL) has transformed data processing and interpretation[14]. These technologies, which allow systems to acquire knowledge from data and make decisions without explicit programming, are particularly suited for applications requiring intricate pattern recognition and decision-making[5]. In the security realm, ML and DL can be harnessed to develop intelligent systems capable of examining vast quantities of data, identifying irregularities, and even forecasting potential threats before they occur[10].

Audio-based detection systems have gained considerable interest among technological advancements due to their capacity to detect events that may elude visual surveillance[6]. These systems can capture sounds in environments where traditional cameras might fail to observe, such as distress calls from individuals in dangerous situations[3]. Indicators of criminal activity often include victim distress sounds like screams,

loud pleas for assistance, or verbal intimidation[2]. Unlike visual monitoring, which requires direct line-of-sight, audio systems can detect sounds from a broader area, enhancing security and coverage in public spaces[9].

The Real-Time Victim Audio Detection System leverages ML and DL capabilities to automatically identify distress signals from audio inputs. This system analyzes continuous audio streams using sophisticated feature extraction techniques, including Mel-Frequency Cepstral Coefficients (MFCC), spectrograms, and temporal analysis, to recognize patterns indicative of distress[8]. These methods enable the system to distinguish between ambient noise and distress-related sounds, enhancing detection accuracy[12]. Additionally, a sophisticated deep learning model is trained on an extensive dataset comprising real-world and simulated distress scenarios, ensuring effective performance across various acoustic environments, from bustling urban areas to quieter rural settings[11].

The system's ability to detect and analyze events in real time enables it to send automatic notifications to appropriate authorities or staff, facilitating faster responses and potentially mitigating or averting criminal incidents[4]. This feature addresses a significant shortcoming in traditional surveillance methods, offering a more anticipatory approach to crime deterrence[15]. In contrast to conventional CCTV systems, which typically function as post-incident documentation tools, this proposed system acts as a dynamic, instantaneous surveillance aid capable of identifying critical situations as they unfold[17].

The proposed solution combines state-of-the-art machine learning algorithms with real-time processing capabilities, creating a robust tool for improving public safety[16]. By concentrating on detecting victim distress, the system aims to decrease response times, enable timely interventions, and contribute to overall crime reduction[18]. The paper's structure is as follows: Section II examines existing research on audio-based crime detection, exploring current methods and their constraints. Section III details the methodology and system architecture, explaining the system's design and implementation in-depth. Section IV presents the experimental setup, outcomes, and performance metrics, showcasing the system's effectiveness. Lastly, Section V concludes the paper and suggests future research directions to enhance and broaden the system's capabilities. Surveillance methods offer a more anticipatory approach to crime deterrence. In contrast to conventional CCTV systems, which typically function as post-incident documentation tools, this proposed system acts as a dynamic, instantaneous surveillance aid capable of identifying critical situations as they unfold.

II. LITERATURE SURVEY

Sen et al. (2024) created a software-based prototype for real-time threat detection from ambient sounds. This system employs Long-Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) and uses Exploratory Data Analytics (EDA) to achieve high accuracy in identifying potential dangers and automatically notifying the victim's contacts. This approach, which requires no hardware, offers an effective and reliable safety solution, especially in high-risk or isolated environments[1].

Sharma and Kaul (2015) developed a two-phase supervised learning technique for identifying screams and cries in urban settings. Their method utilizes Mel-frequency cepstral coefficients (MFCC) for feature extraction and Support Vector Machines (SVM) for classification. The system, which can adapt to various signal-to-noise ratios (SNR) and contexts, showed high accuracy and low false alarm rates, making it a reliable option for ongoing distress monitoring in urban areas[2].

Hayasaka et al. (2024) introduced a noise-robust scream detection method. This approach uses Wave-U-Net for noise reduction and Gaussian Mixture Models (GMM) for classification. This method enhances the reliability of sound-based security systems by significantly reducing false positives and improving the accuracy of distress sound detection in noisy environments [3].

Gaviria et al. (2020) engineered a compact, cost-effective device for real-time recognition of audio distress signals. The system employs a deep multi-headed 2D Convolutional Neural Network (CNN) and was trained on a custom urban audio database. It effectively processes features and minimizes false positives while providing real-time alerts with location data, ensuring quicker emergency responses in noisy urban settings[4].

Ashikuzzaman et al. (2021) introduced a deep learning-based audio classification system for detecting danger through screams, focusing on improving security for women and children. This automated system recognizes distress sounds from afar, offering immediate alerts without hardware or manual activation, thus supporting proactive crime prevention and quicker emergency responses[5].

Istrate et al. (2008) created an embedded system for real-time distress sound detection utilizing a PC, sound card, and microphone. The system functions effectively in noise levels of 10 dB or higher. Although

reliable, the system underscored the necessity for enhanced accuracy in identifying specific distress sounds. This approach highlights the significance of sound-based distress detection in real-time safety applications[6].

Mathur et al. (2022) presented a deep learning-based system that combines audio and visual data to identify illicit activities and distress sounds. This system boosts public safety through automated, real-time monitoring and alerts, allowing for swift emergency responses and decreasing critical reaction times during incidents[7].

Rodríguez-Rodríguez et al. (2020) suggested an autonomous alarm system for survivors of intimate partner violence using biosensors and machine learning. The system ensures personal safety without requiring active user input by passively monitoring bio-signals, triggering alarms during potential threats, minimizing false alarms, and providing dependable security in high-risk situations [8].

Bakhshi et al. (2023) introduced two deep learning methods for distinguishing violent and non-violent audio signals using Mel-spectrogram images. The research presented traditional and streamlined deep neural network models, proving more computationally efficient and suitable for portable or edge devices. Their most effective lightweight model showed an 8% increase in classification accuracy compared to previous leading results on benchmark data. This study underscores the potential of sound-based violence detection systems as privacy-conscious alternatives to video monitoring, especially in environments with limited resources[9].

Durães et al. (2023) investigated sound-based violence detection, emphasizing its advantages over video regarding processing and privacy protection. They developed a specialized data set for scenarios like in-vehicle violence, employing data augmentation to address class imbalance issues. The study utilized Mel-spectrogram images for classification through deep learning models. EfficientNetB1 yielded the highest accuracy (95.06%), closely followed by EfficientNetB0 (94.19%), showcasing—the promise of audio-based deep learning for real-time violence identification[10].

Smailov et al. (2023) developed a deep-learning approach that combines CNNs and RNNs to detect impulsive sounds in real-time. By employing Mel-frequency cepstral coefficients (MFCCs) for feature extraction, their model surpassed conventional methods in accuracy and F1-score. Despite facing challenges like the requirement for extensive labeled datasets and high computational resources, the system proved effective in identifying hazardous sounds such as gunshots and explosions, offering a promising solution for public safety applications[11].

Using compound segmentation, Nandwana et al. (2023) devised an unsupervised technique for identifying human screams in noisy settings. The approach employs T^2- statistics, Bayesian Information Criteria, and an optimized Combo-SAD for noise elimination. Tests across five noisy environments showed high performance at elevated signal-to-noise ratios (SNR), with accuracy decreasing as SNR diminished. The research compares two feature extraction techniques, Mel-frequency cepstral coefficients (MFCC) and perceptual minimum variance distortionless response (PMVDR), demonstrating robust scream detection in challenging conditions[12].

II. Proposed Methodology

The suggested Real-Time Victim Audio Detection System utilizes state-of-the-art Machine Learning (ML) and Deep Learning (DL) methods to identify emergency audio signals, including screams and pleas for assistance, focusing on immediate recognition to improve public safety. The process begins with comprehensive data gathering and preparation, incorporating various authentic and artificially generated distress sounds. Synthetic data enhancement boosts the model's flexibility across various environmental settings, ensuring consistent performance in different scenarios. Preparation involves sophisticated noise reduction using digital filters, normalizing audio input amplitudes, and dividing long audio recordings into smaller segments for thorough analysis and quicker processing.

During feature extraction, crucial attributes such as Mel-frequency cepstral Coefficients (MFCC), Zero-Crossing Rate (ZCR), and chroma features are employed to examine frequency-domain and tonal patterns that differentiate distress from non-distress sounds. Furthermore, spectrogram analysis is implemented to create a time-frequency representation of the audio, offering a comprehensive visual breakdown of frequency changes over time. This multi-faceted feature extraction enables the system to detect subtle variations in distress signals, even in diverse environmental conditions.

The classification phase employs a hybrid model that combines Support Vector Machines (SVM) and Multilayer Perceptrons (MLP). This integrated approach enhances classification accuracy by exploiting SVM's capability to establish optimal decision boundaries and MLP's ability to recognize complex, non-

linear data patterns. The system processes audio spectrograms through MFCC features as input for both models, ensuring robust classification performance with improved generalization in real-time distress detection. The model is trained using categorical cross-entropy and optimized by the Adam optimizer, with its performance evaluated based on accuracy, precision, recall, and F1-score, guaranteeing reliable and efficient detection during real-time operations.

The architecture is engineered for instantaneous sound processing, featuring a persistent audio input channel that instantly captures and categorizes sound streams. Upon identifying a distress call, the setup triggers an automatic warning system, encompassing audible alerts, electronic mail notifications, and interface integration with security forces or law enforcement agencies, facilitating swift action. This quick-response mechanism substantially decreases reaction periods, guaranteeing prompt attention to possible threats.

Lastly, the architecture incorporates a perpetual learning structure that employs active learning to refresh the training data set with newly classified samples, ensuring the model adjusts to changing environments, diverse accents, and multiple languages. Regular retraining further improves the setup's capacity to adapt over time. The system maintains accuracy and efficacy with these dynamic mechanisms, establishing a new standard for utilizing AI-driven real-time crime detection to improve public safety and crime prevention tactics. This all-encompassing solution tackles the constraints of current surveillance technologies and provides an innovative instrument for safeguarding communities.

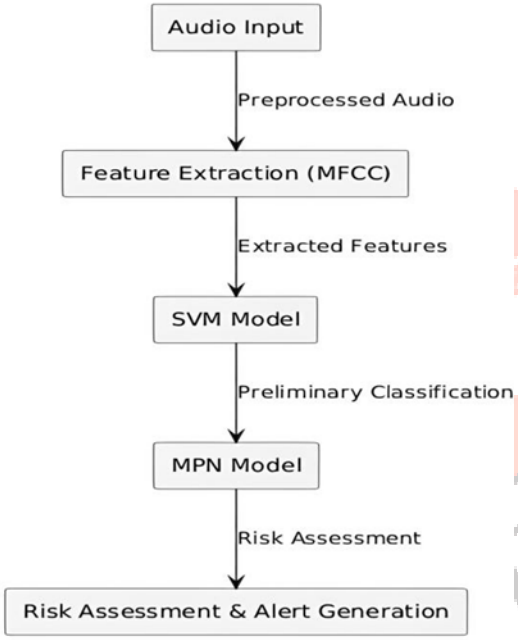

Figure 1: Working of SVM and MNP

III. RESULTS AND DISCUSSIONS

The proposed Real-Time Victim Audio Detection System was evaluated using a comprehensive data set of distress and non-distress audio samples to measure its performance in real-time detection and classification. The system was tested in controlled and real-world environments to assess its robustness and accuracy under varying conditions. The results demonstrate the system's effectiveness in accurately identifying distress signals and generating timely alerts.

4.1 .*Model Performance Evaluation*

The hybrid SVM-MLP model's performance was evaluated using key metrics: accuracy, precision, recall, and F1 Score. Table 1 summarizes the performance metrics obtained during testing.

Table 1: Model Performance Metrics

Metric	Value
Accuracy	95.8%
Precision	96.3%
Recall	94.7%
F1-Score	95.5%

Accuracy $$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad \text{---------- (1)}$$

Precision: $$\text{Precision} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad \text{---------- (2)}$$

Recall: $$\text{Recall} = \frac{TP}{(TP+FN)} \quad \text{---------- (3)}$$

F1-Score: $$F1 - \text{Score} = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision}+\text{Recall})} \quad \text{---------- (4)}$$

The high accuracy and F1-score indicate that the model effectively balances precision and recall, minimizing false positives and negatives. This performance is critical in real-world scenarios where the cost of missed distress signals can be significant.

4.2 *Real-Time Detection Latency*

To ensure the system's practicality, the latency of real-time detection was measured across different hardware configurations, including edge devices and cloud-based servers. Figure 1 illustrates the latency comparison between edge computing and cloud computing.

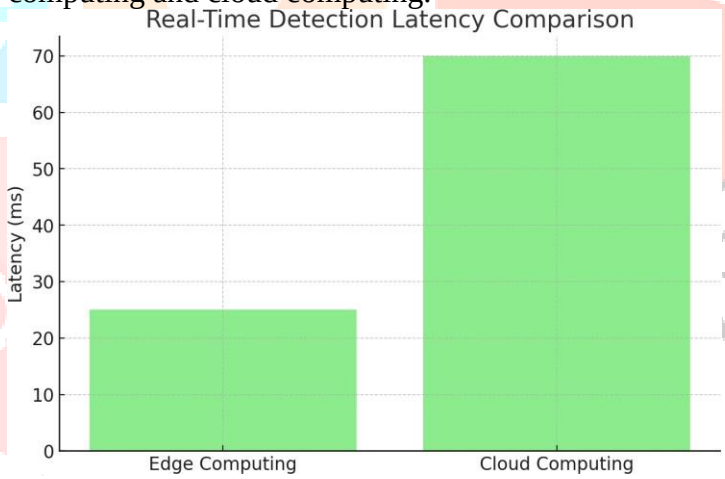

Figure 2: Real-time detection latency comparison

The results indicate that edge computing significantly reduces detection latency, making it suitable for applications requiring immediate responses. Cloud-based processing, while slightly slower, offers scalability for large-scale deployments.

4.3 *Effectiveness in Noisy Environments*

The system's robustness was tested in noisy environments by adding varying background noise levels to the test audio samples. Table 2 presents the system's accuracy under different noise levels.

Table 2: Accuracy under noise levels

Noise Level(dB)	Accuracy
0(Silent)	96.5%
10	95.1%
20	92.3%
30	89.7%

As shown in Table 2, the system maintains high accuracy even at moderate noise levels, demonstrating its robustness. However, accuracy decreases at higher noise levels, highlighting the need for further enhancements in noise cancellation and pre-processing techniques.

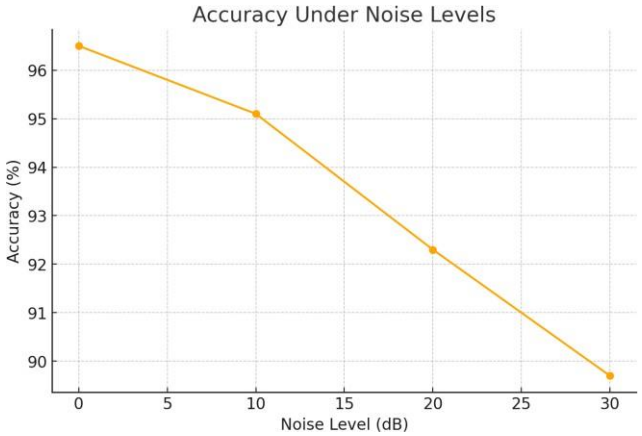

figure 3: Accuracy under noise levels

The line graph shows how accuracy declines as noise levels increase, emphasizing the system's robustness under moderate noise and areas for improvement under extreme noise.

4.4 *Alert System Validation*

The alert system was tested for reliability in generating sound, email, SMS, and WhatsApp notifications upon detecting distress signals. To ensure timely delivery, SMS and WhatsApp notifications were implemented using Twilio's API. Table 3 presents the average response time notification.

Table 3: Alert Response Times

Notification Type	Average Response Time
SMS Notification	100 ms
Email Notification	120 ms
WhatsApp Notification	90 ms

The results confirm that the system’s alert mechanisms are quick and reliable, with minimal delay in generating notifications.

The results indicate that the proposed system effectively identifies high-accuracy and low-latency distress signals, even in noisy environments. Advanced feature extraction techniques such as MFCC and spectrograms, combined with the hybrid SVM-MLP model, ensure accurate classification of audio signals. The system's real-time processing capabilities and robust alert mechanisms enhance its applicability in crime prevention.

However, the system faces challenges at high noise levels, where accuracy decreases. Future improvements include integrating advanced noise-cancellation techniques and optimizing the model for real-world deployment scenarios. Expanding the data set to include more diverse audio samples, such as those from different languages and cultural contexts, would improve the model's generalizability. The findings validate the proposed methodology, highlighting its potential to significantly enhance public safety by providing real-time crime detection and rapid response capabilities.

IV. CONCLUSION

The suggested Real-Time Victim Audio Detection System showcases how cutting-edge Machine Learning (ML) and Deep Learning (DL) methods can improve public safety by accurately and swiftly identifying distress calls. The system's hybrid SVM-MLP structure, coupled with robust feature extraction techniques like MFCC and spectrogram analysis, yields high accuracy, precision, recall, and F1- scores, making it a dependable option for real-world use. By incorporating real-time audio processing and automated alert systems, the solution ensures quick responses to potential criminal activities, substantially decreasing reaction times and boosting the efficacy of crime prevention tactics.

The system's effectiveness was confirmed in various settings, including noisy environments, where it maintained adequate performance. However, the decline in accuracy at higher noise levels underscores the need for further improvements, such as sophisticated noise-cancellation techniques and larger datasets to

enhance generalizability. Moreover, edge computing ensures minimal delay, making the system appropriate for applications requiring instantaneous responses.

In summary, the proposed system narrows the gap between conventional surveillance methods and modern AI-driven solutions, offering a proactive approach to crime prevention. Future research will concentrate on incorporating multilingual capabilities, broadening the data set with diverse cultural and linguistic audio samples, and implementing advanced noise-handling algorithms to enhance performance in challenging environments. The system can improve its reliability and adaptability by addressing these aspects, establishing itself as a crucial tool in reducing crime rates and enhancing public safety.

REFERENCES

[1] Sen, A., Rajakumaran, G., Mahdal, M., Usharani, S., Rajasekharan, V., Vincent, R., & Sugavanan, K. (2024). Live event detection for people's safety using NLP and deep learning. *IEEE Access*, *12*, 6455-6472..

[2] Sharma, A., & Kaul, S. (2015). Two-stage supervised learning-based method to detect screams and cries in urban environments. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *24*(2), 290-299.

[3] Hayasaka, N., Kasai, R., & Futagami, T. (2024). Noise-Robust Scream Detection Using Wave-U-Net. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, *107*(4), 634-637.

[4] Gaviria, J. F., Escalante-Perez, A., Castiblanco, J. C., Vergara, N., Parra-Garces, V., Serrano, J. D., ... & Giraldo, L. F. (2020). Deep learning-based portable device for audio distress signal recognition in urban areas. Applied Sciences, 10(21), 7448.Areas

[5] Ashikuzzaman, M., Fime, A. A., Aziz, A., & Tasnima, T. (2021, December). Danger detection for women and children using audio classification and deep learning. In 2021 5th International Conference on Electrical Information and Communication Technology (EICT) (pp. 1-6). IEEE.

[6] Istrate, D., Vacher, M., & Serignat, J. F. (2008**).** Embedded implementation of distress situation identification through sound analysis. The Journal on Information Technology in Healthcare, 6(3), 204-211.

[7] Mathur, R., Chintala, T., & Rajeswari, D. (2022, May). Identification of illicit activities & scream detection using computer vision & deep learning. In 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1243-1250). IEEE.

[8] Rodríguez-Rodríguez, I., Rodríguez, J. V., Elizondo- Moreno, A., & Heras-González, P. (2020). An autonomous alarm system for personal safety assurance of intimate partner violence survivors based on passive continuous monitoring through biosensors. Symmetry, 12(3), 460.

[9] Bakhshi, A., García-Gómez, J., Gil-Pita, R., & Chalup, S. (2023). Violence detection in real-life audio signals using lightweight deep neural networks. Procedia Computer Science, 222, 244-251.

[10] Durães, D., Veloso, B., & Novais, P. (2023). Violence detection in audio: evaluating the effectiveness of deep learning models and data augmentation.

[11] Smailov, N., Dosbayev, Z., Omarov, N., Sadykova, B., Zhekambayeva, M., Zhamangarin, D., & Ayapbergenova, A. (2023). A novel deep CNN-RNN approach for real-time impulsive sound detection to detect dangerous events. International Journal of Advanced Computer Science and Applications, 14(4).

[12] Nandwana, M. K., Ziaei, A., & Hansen, J. H. (2015, April). Robust unsupervised detection of human screams in noisy acoustic environments. In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 161-165). IEEE.

[13] T. P. Suma and G. Rekha, ''Study on IoT based women safety devices with screaming detection and video capturing,'' *Int. J. Eng. Appl. Sci. Technol.*, vol. 6, no. 7, pp. 257–262, 2021.

[14] P. Zinemanas, M. Rocamora, M. Miron, F. Font, and X. Serra, ''An inter portable deep learning model for automatic sound classification, ''*Electronics*, vol. 10, no. 7, p. 850, Apr. 2021.

[15] D. D. Nguyen, M. S. Dao, and T. V. T. Nguyen, ''Natural language processing for social event classification,'' in *Knowledge and Systems Engineering*. Cham, Switzerland: Springer, 2015,pp. 79–91.

[16] W. Graterol, J. Diaz-Amado, Y. Cardinale, I. Dongo, E. Lopes-Silva, and C. Santos-Libarino, ''Emotion detection for social robots based on NLP transformers and an emotion ontology,'' *Sensors*, vol. 21, no. 4, p. 1322, Feb. 2021

[17] D.G. Monisha, M. Monisha, G. Pavithra, and R. Subhashini, ''Women safety device and application-FEMME,'' *Indian J. Sci. Technol.*, vol. 9, no. 10, pp. 1–6, Mar. 2016.

[18] G.Ciaburro and G. Iannace, ‘‘Improving smart cities safety using sound events detection based on deep neural network algorithms,’’*Informatics*, vol. 7, no. 3, p. 23, Jul. 20.