www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 10, October 2025, pp : 893-902

e-ISSN : 2583-1062
Impact Factor : 7.001

CRIME PREVENTION THROUGH SCREAM DETECTION

Sudharsan G[1], Kishore VJ[2], Raghul Sekar[3], Kalpana R[4]

[1,2,3]Student, Department of CSE AI & DS, VELS University, Chennai, India.

[4]Assistant Professor, Department of CSE AI & DS, VELS University, Chennai, India.

Email: sudharsang0714@gmail.com

DOI: https://www.doi.org/10.58257/IJPREMS44170

ABSTRACT

The increasing rate of crimes, especially in urban areas, calls for the development of advanced technologies to ensure public safety and timely intervention. Crime Prevention through Scream Detection, presents an Al-powered system designed to detect human screams in real-time using audio analysis techniques. The primary objective is to identify distress situations, such as assaults or emergencies, where a scream is likely to occur, and automatically trigger alerts to nearby authorities or security systems. The system employs machine learning models trained on a dataset containing both positive (scream) and negative (non-scream) audio samples. By extracting relevant features such as pitch, frequency, and amplitude, the model effectively differentiates between normal environmental sounds and actual screams. Once a scream is detected, the system can be configured to activate alarms, notify law enforcement, or record the event for further investigation. The system can be integrated with smart surveillance cameras, home security systems, and mobile applications, making it versatile for deployment in residential areas, public spaces, transportation networks, and institutional environments. Real-time detection enables faster response and timely intervention, thereby preventing crimes or minimizing their severity.

Keywords: Crime Prevention, Scream Detection, Audio Analysis, Machine Learning, Real-time Detection, Public Safety.

1. INTRODUCTION

Crime prevention remains one of the foremost concerns in maintaining public safety and security, particularly in densely populated urban areas. With the rapid pace of urbanization, incidents such as assault, theft, and harassment are becoming increasingly frequent. Unfortunately, in many such cases, victims find themselves unable to immediately seek help due to fear, physical restraint, or lack of access to communication devices. Nevertheless, one of the most natural and instinctive responses in dangerous situations is to scream or produce distress sounds. These sounds serve as immediate signals of danger, alerting nearby individuals or bystanders to potential threats. Leveraging this instinctive behaviour through technology offers an innovative way to improve crime prevention and response.

This project focuses on developing an intelligent system capable of automatically detecting screams using advanced audio signal processing and machine learning techniques. Unlike traditional surveillance systems that primarily rely on video feeds, the proposed system utilizes the auditory dimension of human behaviour, which can be especially useful in situations where visual data is either unavailable or unclear. For example, crimes or emergencies that occur in poorly lit areas, crowded spaces, or behind obstacles may be missed by cameras. However, the sound of a scream or distress call can still be captured and analysed, making this auditory approach a highly effective method for detecting danger in real-time.

The proposed system goes beyond scream detection by integrating real-time audio analysis with automated alert mechanisms. These mechanisms are designed to notify relevant authorities, security personnel, or nearby individuals whenever a potential threat is identified. By minimizing the delay between the occurrence of an incident and the initiation of a response, the system significantly enhances safety outcomes. Furthermore, it reduces the dependency on continuous manual monitoring of video or audio feeds, which is often resource-intensive and susceptible to human error.

This technology aligns well with the vision of smart cities, where advanced systems are integrated into public infrastructure to enhance safety and improve quality of life. By incorporating scream detection into intelligent surveillance networks, cities can develop more responsive and efficient systems that protect their residents. Additionally, the system's ability to detect distress signals across various environments makes it a versatile tool for enhancing public safety.

From a technical standpoint, the system leverages audio signal processing techniques to extract unique features from screams that distinguish them from other sounds, such as normal speech, laughter, or background noise. These features are then fed into machine learning models, which have been trained to identify and classify screams in real time. The challenge here lies in handling noisy environments, differentiating between false alarms, and ensuring the system's reliability and accuracy.

In real-world applications, one of the key challenges to address is minimizing false positives. Scream detection systems must be able to distinguish between distress signals and everyday noises, such as loud conversations, laughter, or traffic sounds. For the system to be practical and effective, it must function reliably across a range of environments, from quiet residential streets to bustling urban centres. Additionally, privacy concerns need to be carefully considered, as the system would rely on constant audio monitoring. To mitigate these concerns, privacy-enhancing features—such as anonymization of the audio data and ensuring that only relevant alerts are recorded—would be integrated into the system design.

One of the most significant benefits of this technology is its ability to drastically reduce response times during emergencies. By alerting security personnel, law enforcement, or even nearby citizens the moment a scream is detected, the system enables rapid intervention. This faster response can potentially prevent a crime from escalating and save lives. Furthermore, by providing a real-time, automated alert mechanism, the system frees up human resources, allowing security personnel to focus on addressing the threat rather than constantly monitoring surveillance feeds.

This project presents an innovative approach to crime prevention and public safety by leveraging the natural human reaction of screaming to trigger automated responses. The intelligent system developed is capable of detecting screams in real time and initiating alerts to ensure that safety measures are enacted without delay. Its applicability across diverse environments—ranging from public spaces to transportation systems—highlights its versatility and relevance in enhancing security. This technological framework offers significant potential for future crime prevention strategies, contributing to the advancement of smart, responsive, and efficient safety infrastructures. By integrating advanced audio analytics with machine learning, the project provides a novel perspective on strengthening public security and addressing emerging challenges in crime prevention.

2. LITERATURE REVIEW

Scream and shout detection has gained significant attention in recent years due to its potential applications in public safety, surveillance, and emergency response systems. Traditional sound recognition techniques typically focus on speech, music, or environmental sounds, but detecting screams presents unique challenges. These challenges arise from the variability in human screams, the wide range of environmental conditions in which they occur, and the difficulty of distinguishing distress sounds from background noise or other loud events. As such, previous research has focused on different approaches, particularly in the areas of acoustic feature extraction, signal processing, and machine learning algorithms, to improve the accuracy and efficiency of scream detection systems.

One major research direction has concentrated on differentiating between screams and other high-intensity sounds, especially in noisy environments. For example, a study [1] used microphone arrays in combination with Time Difference of Arrival (TDOA) techniques to distinguish screams from gunshots. The researchers employed two parallel Gaussian Mixture Models (GMMs) to model the different sound events, achieving an impressive detection accuracy of 93% and a 5% false alarm rate. This study demonstrated the potential of using probabilistic models for sound classification, particularly in environments with multiple competing sounds. Similarly, another study [2] addressed the problem of sound recognition in subway systems, where background noise levels are typically high. This research explored the recognition of screams, explosions, and gunshots, emphasizing the need for robust feature extraction and classification methods that could perform well in real-world, noisy environments.

In addition to scream detection in specific scenarios, other research has tackled broader sound event classification, with screams being part of a larger dataset. One such study [3] developed a model capable of recognizing a variety of everyday sounds, such as glass breaking, water flowing, and screams. While the system was not tailored specifically for distress sound detection, it provided valuable insights into the challenges of sound classification in noisy environments. Another study [4] tested detection algorithms on a diverse dataset containing sounds like claps, laughter, knocks, and explosions, aiming to evaluate performance under different acoustic conditions. While these studies helped advance the general field of sound classification, they often lacked specific focus on the nuances of scream detection and the need for fine-tuned models to accurately identify distress signals.

Several techniques for feature extraction and classification have been explored to improve scream detection accuracy. One study [5] combined acoustic features with Support Vector Machine (SVM) classifiers to identify sound events, achieving promising results in differentiating between various types of sounds. Another study [6] utilized Mel-Frequency Cepstral Coefficients (MFCCs) and MPEG-7 audio descriptors, along with Hidden Markov Models (HMMs), to classify gunshots and screams. These approaches demonstrated the importance of using a combination of feature extraction methods and classifiers to improve performance. Research [7] further expanded on this concept by integrating multiple GMM classifiers for different sound categories, achieving an overall accuracy of 90% with an 8% false alarm rate. Despite these advances, minimizing false positives in complex, real-world acoustic environments remain an ongoing challenge.

Distinguishing between shouted speech and pure screams has also been a topic of investigation. Studies [8] and [9] analysed the phonetic and acoustic characteristics of shouted words, focusing on how prosodic features (e.g., pitch, intensity, and duration) and energy levels differ between loud speech and distress signals. These findings have informed the development of more refined models capable of differentiating between normal loud speech and screams, addressing the overlap between these two types of vocalizations.

Our approach builds on the insights from previous studies but introduces several innovations. Unlike earlier methods that primarily relied on statistical techniques such as SVMs, GMMs, or T2-statistics [1], we employ a unique three-phase classification process. In the first phase, environmental noise is filtered to enhance the signal quality and remove unwanted background sounds. The second phase focuses on isolating speech from non-speech events, which is essential for distinguishing screams from other high-intensity noises. The final phase is specifically dedicated to detecting and classifying screams. To improve detection accuracy, we integrate a Support vector machine-based classifier with a Multilayer Perceptron (MLP) neural network. This hybrid approach combines the strengths of both models, enhancing the system's ability to distinguish between screams, shouts, regular speech, and background noise.

Another key differentiator in our approach is the use of real-life scream recordings for training the model. While many previous studies relied heavily on simulated or synthetic datasets, our model is trained on real-world scream data collected from volunteers. This ensures that the system is exposed to the natural variability in human vocal expressions and diverse environmental conditions that can occur in real-world situations. By training on actual scream recordings, we improve the model's ability to generalize across different scenarios, making it more applicable to practical applications in public safety and surveillance systems.

3. METHODOLOGY

The proposed system for automatic scream detection is designed as a multi-stage pipeline that integrates several key components: data collection, feature extraction, machine learning and deep learning classification, deployment, and real-time alerting. Each stage in the process is essential to ensure that the system operates accurately, reliably, and effectively in real-world settings. The methodology can be divided into five major phases: Audio Data Collection, Preprocessing and Feature Extraction, Model Training, Deployment and Integration, and the Alert Mechanism.

3.1 Audio Data Collection

The first step in the methodology is the collection of a comprehensive dataset consisting of both scream and non-scream sounds. For the system to be effective in a wide range of real-world situations, it is essential that the dataset captures a broad spectrum of scream variations. These include differences in intensity, pitch, emotional tone, and duration. To ensure diversity, scream recordings are gathered from volunteers under controlled conditions, which allows for a varied set of scream types, including both natural and exaggerated distress calls. In addition, publicly available audio databases containing real-life sound events are used to supplement the dataset.

Non-scream sounds, such as speech, laughter, applause, and various environmental noises, are also included to ensure that the model can distinguish between screams and other loud events. The variety in the dataset, which encompasses both scream and non-scream categories, helps to increase the robustness and generalization of the detection system, ensuring that it can perform well across different environments and scenarios.

3.2 Preprocessing and Feature Extraction

After the raw audio data is collected, it is subjected to preprocessing to improve the quality and consistency of the recordings. The preprocessing steps include noise reduction, where background noise is minimized to ensure clearer sound signals, and normalization of sound levels to reduce inconsistencies between recordings. Additionally, long audio recordings are segmented into shorter clips, which facilitates more manageable processing and analysis.

The key step in preprocessing is feature extraction, where relevant acoustic features are derived from the raw audio signals. For this project, Mel-Frequency Cepstral Coefficients (MFCCs) are selected as the primary feature extraction method. MFCCs are widely used in speech and sound recognition tasks because they effectively capture the spectral and temporal properties of human voice signals. Specifically, MFCCs provide a representation of the frequency content of the sound, which is particularly useful for distinguishing the distinct characteristics of screams from other types of audios, such as speech or environmental noise. MFCCs are also relatively robust to variations in background noise, making them suitable for deployment in noisy environments.

3.3 Model Training

The heart of the system lies in training machine learning and deep learning models that can classify scream and non-scream sounds. Two distinct approaches are employed to optimize performance and compare the effectiveness of traditional machine learning and deep learning techniques:

Support Vector Machine (SVM): This is a supervised machine learning algorithm commonly used for binary classification tasks. In this project, the SVM model is trained using the extracted MFCC features to create a decision boundary that separates scream samples from non-scream samples. The SVM algorithm is particularly effective in high-dimensional spaces, making it suitable for the complex feature sets derived from audio signals.

Multilayer Perceptron (MLP): This is a deep learning model that uses multiple layers of neurons to capture non-linear relationships in the data. The MLP is trained on the same MFCC features, allowing the model to learn more complex representations of scream characteristics, such as emotional tone, intensity, and frequency patterns. By leveraging its deep network structure, the MLP is capable of learning higher-level patterns and subtle distinctions between scream sounds and other non-scream events.

By comparing the performance of both models, the system aims to combine the strengths of traditional machine learning and deep learning approaches. Key performance metrics, including accuracy, precision, recall, and false alarm rates, are used to evaluate and compare the effectiveness of the models. The best-performing model is then selected for deployment.

3.4 Deployment and Integration

Once the models are trained and optimized, they are converted into lightweight, deployable formats that can be used in real-time applications. This phase involves integrating the trained models into a detection framework that can continuously process incoming audio streams from surveillance systems or embedded devices. The system is designed to analyse live audio in real time, enabling it to detect screams as soon as they occur. The integration is tailored for various environments, ensuring that the system can function effectively in public spaces, transportation systems, or other safety-critical settings.

This phase also includes optimizing the system for efficient processing, ensuring that the scream detection algorithm can run on limited computational resources without compromising detection accuracy. The goal is to make the system practical for widespread deployment, including in situations with constrained hardware, such as low-cost surveillance cameras or wearable devices.

3.5 Alert Mechanism

The final phase of the methodology is the implementation of the alert mechanism. Once the system detects a scream with high confidence, it triggers an automated alert to notify relevant authorities or individuals. The alert system sends an SMS message to a registered phone number, providing information about the detected scream and its location. This location is obtained via GPS integration, which ensures that the alert includes precise details about the geographic coordinates of the incident.

The goal of this alert mechanism is to enable a rapid response to potential dangers, whether from security personnel, emergency responders, or nearby bystanders. By reducing the time, it takes for authorities to become aware of a distress signal, the system helps mitigate potential harm and improve overall public safety. This real-time alerting system is designed to be fast, reliable, and easy to integrate into existing emergency response frameworks.

4. IMPLEMENTATION

The implementation of the scream detection system follows a systematic approach that involves several critical stages: data collection and preprocessing, model training, and the integration of detection mechanisms with an alert system. Each stage is designed to ensure that the system can effectively identify screams and respond in real-time during emergencies.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 10, October 2025, pp : 893-902
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

Figure 4.1 - Architecture

4.1 Data Collection and Preprocessing

The first stage in building the system is collecting and preprocessing the audio data. A reliable scream detection system requires a comprehensive dataset that includes both scream sounds and normal sounds. Normal sounds consist of everyday audios like speech, laughter, background noise (e.g., traffic, birds, crowds), and other environmental sounds commonly found in public and private spaces. The scream data, on the other hand, can be sourced from public audio databases, controlled experimental recordings, or extracted from films and online media. It's essential to ensure that the scream dataset is diverse, encompassing variations in gender, age, loudness, pitch, and emotional intensity. Human screams can vary significantly depending on individual characteristics and situational factors, so diversity is key to improving the system's accuracy across different contexts.

Once the dataset is collected, the next step is preprocessing the raw audio to prepare it for analysis. Preprocessing includes several tasks:

1. Noise Reduction: This step eliminates unwanted background interference, such as hums or irrelevant sounds, which could obscure the primary signal.

2. Silence Removal: Silent sections of the recordings are removed to focus only on the parts that contain meaningful sound.

3. Normalization: Audio clips are normalized to a standard volume level, ensuring that differences in microphone sensitivity or recording conditions don't affect the model's ability to learn.

4. Segmentation: Long audio recordings are divided into shorter, manageable clips. This segmentation allows the system to process and analyse real-time audio in smaller chunks, ensuring faster and more efficient detection.

5. After preprocessing, feature extraction takes place. This is where key acoustic features are derived from the raw audio, transforming it into numerical data that can be used for classification. Commonly used features include:

6. Mel-Frequency Cepstral Coefficients (MFCCs): MFCCs are effective at capturing the spectral and temporal properties of human vocal signals, which makes them particularly useful for distinguishing screams from other sounds.

These features are then used for training machine learning or deep learning models.

4.2 Model Training

The next phase focuses on training models for scream classification. In this step, we use a combination of traditional machine learning algorithms and deep learning models to classify audio clips as either scream or non-scream sounds.

Traditional Machine Learning Model: Support Vector Machines (SVM) is used with the extracted features to classify audio. SVM is effective when the feature space is relatively low-dimensional and a clear decision boundary exists between categories, making it well-suited for audio classification tasks.

Deep Learning Model: Multilayer Perceptron’s (MLPs) are used for audio classification by taking extracted features as input. MLPs consist of multiple fully connected layers that can learn complex, non-linear relationships in the data.

They are effective when the input features capture relevant characteristics of the audio, enabling the model to distinguish between different sound categories. MLPs can capture patterns in the feature space, making them suitable for detecting screams in audio signals.

During the training process, the dataset is split into training, validation, and test sets. This allows us to evaluate the model's accuracy, precision, recall, and F1-score—metrics that are crucial for assessing the model's performance. Techniques such as data augmentation, dropout, and hyperparameter tuning are employed to prevent overfitting, improve robustness, and optimize the model's performance.

4.3 Real-Time Detection and Alert Integration

Once the model is trained and optimized, it is integrated into a real-time detection framework. This framework continuously monitors incoming audio streams from microphones or surveillance systems. As audio is analysed, the model classifies it in real-time and triggers an automatic response if a scream is detected.

The trained model is deployed in a lightweight format that can run on edge devices like smartphones. This ensures the system operates efficiently, even in environments with limited computational resources. Both the model predicts

High Risk and one of the model predicts – Medium Risk. No model Predicts – No Risk.

When a scream is detected, the system automatically triggers an alert. This alert can take various forms, such as sending an SMS or notification to emergency contacts or local authorities. Additionally, the alert system can be integrated with GPS modules, allowing it to provide precise location data to responders. This enables a faster and more accurate emergency response, as the authorities are guided directly to the location of the event.

To minimize false alarms, the system allows for threshold settings. These thresholds ensure that an alert is only triggered when the model's confidence in detecting a scream is above a certain level. For instance, if the model identifies a sound with 80% confidence as a scream, but the threshold is set to 90%, no alert will be sent until the detection surpasses the threshold. This helps ensure that only genuine scream events trigger notifications, reducing unnecessary alerts.

5. RESULTS AND DISCUSSION

The proposed scream detection system aims to improve personal safety and enable faster intervention during emergencies. By focusing on three core outcomes—functioning detection, automatic alert generation with location details, and quicker crime response—the system is designed to be practical, reliable, and impactful in real-world scenarios. Each of these outcomes plays a critical role in enhancing public and private safety by ensuring timely alerts and interventions in distress situations.

5.1 Development of a Functional Detection System

The first objective of the scream detection system is to develop an effective mechanism for identifying human screams. Screams are unique in their acoustic properties, including pitch, intensity, and temporal variations, which make them distinct from regular speech or environmental noise. Human screams typically have a higher frequency and louder intensity compared to other sounds, which makes them identifiable even in noisy environments.

The system employs advanced audio signal processing techniques to extract meaningful features from audio signals, such as Mel-Frequency Cepstral Coefficients (MFCCs), spectral centroid, and zero-crossing rate. MFCCs, for instance, are widely used in speech recognition because they capture the frequency characteristics of sound. For scream detection, these features help in distinguishing between distress signals and normal sounds like speech or ambient noise. By utilizing machine learning and deep learning models, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), the system can effectively learn patterns from both the frequency and temporal aspects of audio signals. CNNs excel at analysing spectrograms, which represent audio data visually, while RNNs and Long Short-Term Memory (LSTM) networks are well-suited for handling the sequential nature of audio.

Training the models with a large and diverse dataset ensures that the system can generalize well to different environments and accurately recognize screams. By balancing between accuracy, precision, recall, and minimizing false positives, the system becomes reliable in real-world settings where background noise is often prevalent. The ultimate goal of this phase is to develop a robust detection system that not only identifies screams but does so with a low rate of false alarms, ensuring high sensitivity to genuine distress calls.

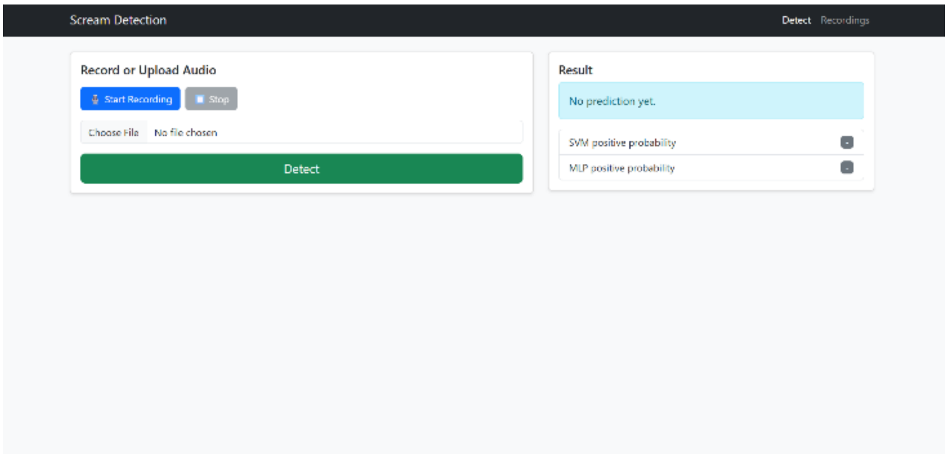

Figure 5.1- Main Page

5.2 Automatic Alert Generation with Location Details

When the system detects a scream classified as medium or high risk, it automatically triggers an alert and shares the user's real-time location with the registered emergency contact. This integration of risk assessment with location tracking ensures that assistance can be dispatched immediately to the precise location, enhancing safety and response efficiency.

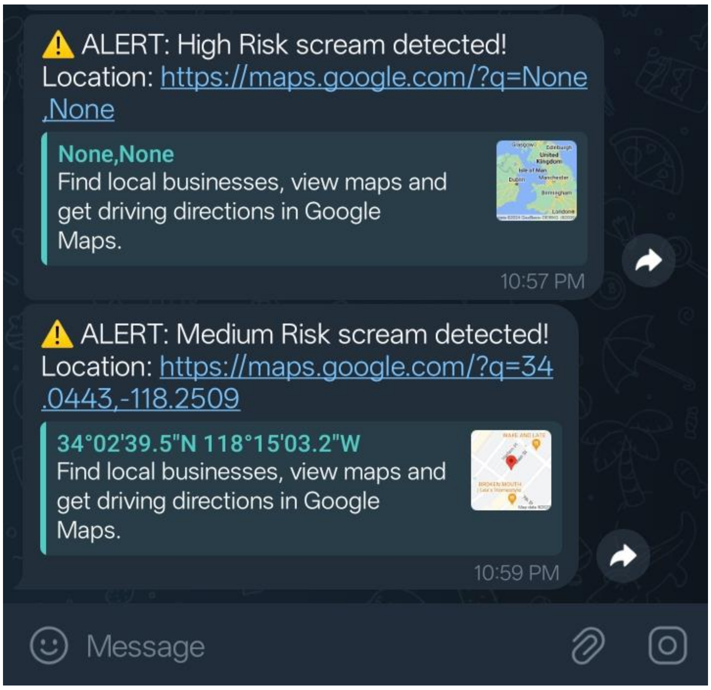

Figure 5.2 - Message

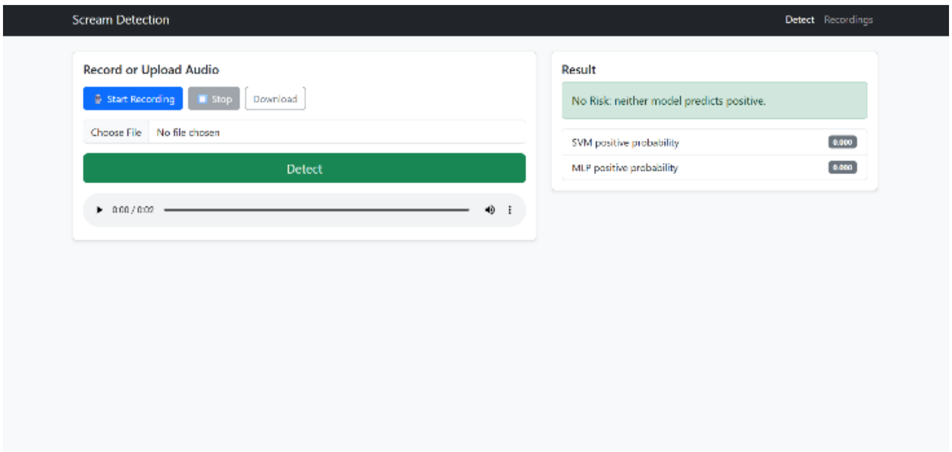

Figure 5.3 - No Risk

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 10, October 2025, pp : 893-902
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

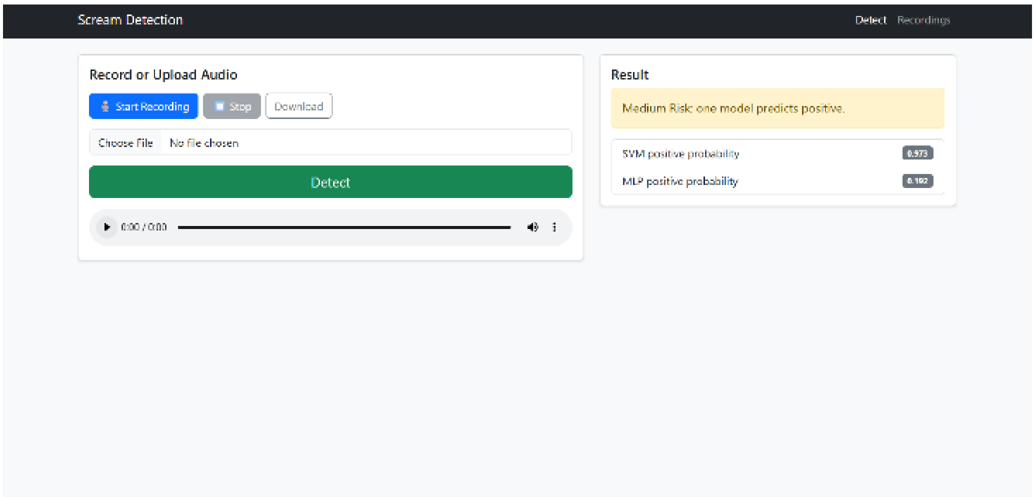

Figure 5.4 - Medium Risk

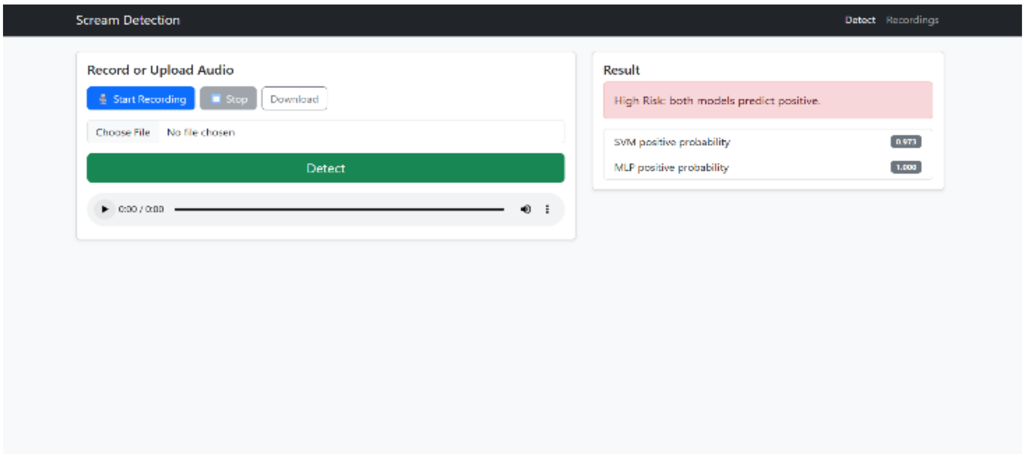

Figure 5.5 - High Risk

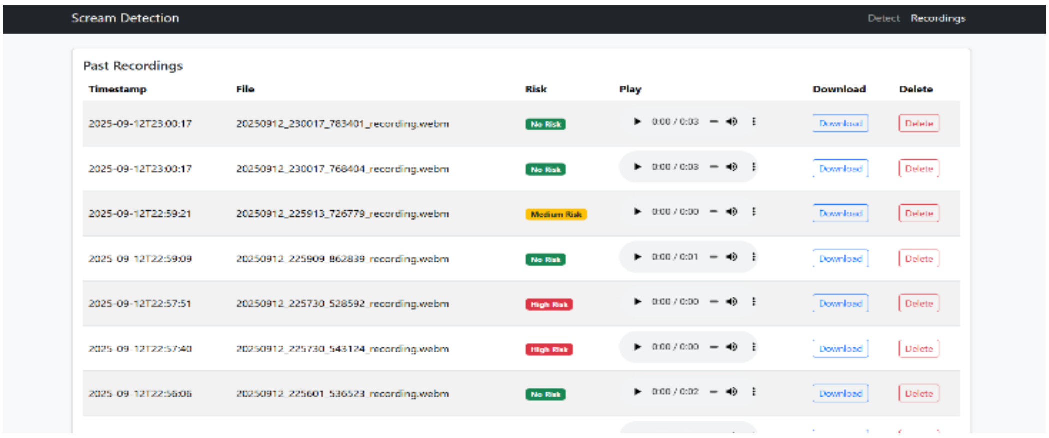

Figure 5.6 - Recordings

5.3 Faster Crime Response and Prevention

The third outcome of the scream detection system is its potential to improve crime response times and prevent crimes by allowing quicker intervention. In many cases, crimes such as assaults, harassment, or even kidnappings happen in situations where the victim is isolated or unable to call for help. For example, in public spaces like parks, public transportation systems, or even in private homes, the victim may be alone or unable to communicate their distress due to physical restraint, fear, or other circumstances.

By deploying the scream detection system in areas where these crimes are most likely to occur—such as public spaces, workplaces, or homes—the system provides a proactive layer of security.

The system's ability to detect screams in real-time acts as a deterrent to potential offenders. Knowing that such a detection system is in place can discourage criminal behaviour, as offenders may recognize that their actions are more likely to be reported and responded to swiftly. Moreover, the widespread implementation of this system could contribute to reducing crime rates and fostering a greater sense of public safety. In addition to protecting individuals, the system could also enhance the overall sense of security within communities. When people feel safe in their surroundings, they are more likely to engage in public activities, which can have positive social and economic impacts. By improving the response times of law enforcement and creating an environment where individuals feel protected, the system plays a role in building safer communities.

6. CONCLUSION

This study demonstrates the potential of scream detection as an effective tool for crime prevention and public safety enhancement. By leveraging audio-based analysis to automatically identify distress signals in real time, the proposed system provides an early warning mechanism that can facilitate rapid intervention and response. The results highlight that scream detection can serve as a valuable complement to existing surveillance and security measures, particularly in environments where visual monitoring alone may be limited. While the findings underscore the promise of this approach, future work should address challenges such as reducing false positives, improving robustness in noisy environments, and integrating the system seamlessly with law enforcement and emergency response protocols.

In the future, the system can be improved by using larger and more diverse sound datasets, combining audio with video feeds for better accuracy, and integrating it with smart city applications. With these improvements, scream detection can play a major role in creating safer environments and ensuring quicker action when crimes or emergencies occur.

7. FUTURE WORK

The scream detection system, designed to improve safety and enhance emergency response, has significant potential for expansion and integration with other technologies to further enhance its utility. As technology continues to evolve and the need for more sophisticated and responsive security systems grows, the system can be adapted and scaled for a wider range of applications. Below are some key directions for future work in this area, focusing on enhanced surveillance capabilities, improved response times, scalability to larger environments, and integration with mobile technologies to increase accessibility and efficiency.

A. Integration with CCTV and IoT Devices for Enhanced Surveillance

One of the most promising areas for future work is the integration of the scream detection system with CCTV (Closed-Circuit Television) and IoT (Internet of Things) devices. Surveillance systems based on cameras are widely used in public spaces and private facilities to monitor and record activity. However, cameras have limitations, especially in situations where visibility is poor (such as in low-light environments or obstructed areas) or in dense, crowded spaces. The scream detection system can fill this gap by providing an additional layer of security, responding to audio cues that may not be captured visually.

By linking the scream detection system with CCTV cameras, the system could not only detect screams but also activate the corresponding camera feeds to focus on the area where the distress signal originated. This would enable security personnel or local authorities to quickly assess the situation through both audio and visual information, improving their ability to respond effectively. Furthermore, integration with IoT devices such as smart doorbells, wearables, or other connected devices could enhance situational awareness. For example, a connected wearable device on a person in distress could trigger the system to detect a scream and immediately send an alert along with location data, offering more precise insights into the event.

B. Connecting with Police Emergency Alert Systems for Quick Response

Another significant enhancement would be integrating the scream detection system with police emergency alert systems. In many cases, emergency situations are not immediately recognized by victims or bystanders, and delayed responses can lead to worsened outcomes. By integrating the system with emergency services, the detection of a scream would trigger an immediate alert to the local police, ambulance, or fire department. This integration would allow emergency responders to receive precise location details along with the nature of the distress (such as a possible assault, medical emergency, or accident), facilitating faster and more accurate responses.

The real-time nature of the system would be vital for reducing response times. For example, if a scream is detected in a public place like a mall or train station, emergency services would be able to get to the exact location without relying on manually made phone calls or alerts from onlookers. This would significantly improve the efficiency of emergency response, increase the chances of preventing or minimizing harm, and provide immediate assistance to those in need.

C. Mobile Application Integration for Nearby Security or Authorities

Mobile application integration is another potential area for future work, which would greatly enhance the system's accessibility and response capabilities. By developing a mobile app connected to the scream detection system, local

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 10, October 2025, pp : 893-902
e-ISSN : 2583-1062
Impact Factor : 7.001

security personnel, authorities, or even nearby individuals could receive real-time alerts when a scream is detected in their vicinity.

For example, if a scream is detected near a building, a registered security officer using the mobile app would receive an alert on their smartphone, which includes the location of the incident, and they could respond accordingly. Additionally, the mobile app could allow the general public to register as responders or notify others nearby, thereby creating a community-based emergency response system. This would be particularly useful in areas that may not have immediate access to formal security resources, but where individuals could help by intervening or guiding authorities to the scene.

Furthermore, the mobile app could feature GPS-based tracking, enabling responders to be guided directly to the location of the distress signal, reducing response time and improving the chances of preventing harm.

8. REFERENCES

[1] J. Benesty, M. M. Sondhi, and Y. Huang, Springer Handbook of Speech Processing. Berlin, Germany: Springer, 2008.

[2] L. Rabiner and B.-H. Juang, Fundamentals of Speech Recognition. Englewood Cliffs, NJ, USA: Prentice Hall, 1993.

[3] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[4] C. Clavel, T. Ehrette, and G. Richard, “Events detection for an audio-based surveillance system,” Proc. IEEE Int. Conf. Multimedia and Expo (ICME), pp. 1306–1309, 2005.

[5] J. Salamon, C. Jacoby, and J. P. Bello, “A dataset and taxonomy for urban sound research,” Proc. ACM Multimedia Conf. (MM), pp. 1041–1044, 2014.

[6] D. Stowell, D. Giannoulis, E. Benetos, M. Lagrange, and M. D. Plumbley, “Detection and classification of acoustic scenes and events,” IEEE Trans. Multimedia, vol. 17, no. 10, pp. 1733–1746, 2015.

[7] K. Lee, “Real-time scream detection for smart surveillance,” Int. J. Security and Its Applications, vol. 9, no. 6, pp. 239–244, 2015.

[8] J. Salamon and J. P. Bello, “Deep convolutional neural networks and data augmentation for environmental sound classification,” IEEE Signal Process. Lett., vol. 24, no. 3, pp. 279–283, 2017.

[9] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.

[10] M. Abadi et al., “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015.

[11] T. L. A. A. W. O. Anwar, F. M. Ghani, and M. S. H. M. Ibrahim, "A survey on sound event detection and classification," IEEE Access, vol. 7, pp. 143023-143035, 2019.

[12] M. H. Y. M. V. S. R. Ravi, “Scream detection and classification using deep learning models for emergency detection,” Proc. IEEE Global Conf. Signal and Info. Processing, pp. 81–85, 2021.

[13] D. P. W. D. M. C. A. Lee, K. Y. H. Liu, and T. J. Kuo, "Speech and noise detection for sound surveillance systems," IEEE Trans. Audio, Speech, and Language Processing, vol. 16, no. 1, pp. 32–43, 2008.

[14] M. J. C. R. R. S. V. L. B. L. Liu, "Real-time event detection and classification in urban sound environments using convolutional neural networks," Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), pp. 3434-3438, 2018.

[15] A. H. A. B. R. B. J. S. H. F. D. Abdi, "Multimodal crime event detection using acoustic sensors," IEEE Trans. Signal Processing, vol. 65, no. 8, pp. 2002–2014, 2017.

[16] P. K. B. P. S. G. A. S. P. A. Koedinger, "Real-time event detection and classification with multimodal surveillance systems," International Journal of Computer Vision and Image Processing, vol. 15, no. 3, pp. 14-30, 2016.

[17] S. J. D. M. N. B. K. L. Jain, "Machine learning for acoustic event detection: A survey," Journal of Signal Processing Systems, vol. 80, pp. 1-18, 2015.

[18] H. C. J. S. S. E. G. B. O. E. J. K. Lee, "Detection of distress cries using deep learning techniques in real-time surveillance," IEEE Trans. Neural Networks and Learning Systems, vol. 29, no. 5, pp. 1341-1349, 2018.

[19] T. K. D. S. A. G. H. S. W. M. Ko, "Sound event detection using spectrogram-based features and SVM classifiers," IEEE Trans. Audio, Speech, and Language Processing, vol. 22, no. 4, pp. 950-961, 2014.

[20] J. M. F. D. R. S. A. M. M. Benetos, "Environmental sound classification with long short-term memory networks," IEEE Signal Processing Letters, vol. 24, no. 7, pp. 1057–1061, 2017.