World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Review Article)

Check for updates

AI-based human scream detection for crime prevention

Rupali Suresh Bhad [1,*] and Harsha R. Vyawahare [2]

[1] *Research Scholar, MTech Computer Science and Engineering, SIPNA College of Engineering and Technology, Amravati.*
[2] *Professor, Computer Science and Engineering, SIPNA College of Engineering and Technology, Amravati.*

World Journal of Advanced Research and Reviews, 2025, 28(02), 2378-2387

Publication history: Received 16 October 2025; revised on 22 November 2025; accepted on 24 November 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.28.2.3887

Abstract

The Human Scream Detection System is a next-generation safety intelligence framework, which, when completed, will autonomously identify and react to expressions of distress, in particular screams, from within live audio streams. Screams are universal acoustics, linked to fear, pain, or perceived threat, which are often expressed by a human being in any language and culture. For this reason, screams have great potential as a means through which an automated threat might be detected. This system combines high-level audio signal processing with machine learning and deep learning models to separate high-intensity human vocalizations from background noise. The main modules critical to its functioning are preprocessing robust to noise, spectral-temporal feature extraction using MFCC and pitch contours, and classification with SVM, MLP, and CNN models. On the autonomous detection of a scream event, this system can automatically raise emergency alerts and send geolocation information to nearby responding units through secure APIs like Twilio. Other areas of application include smart home security, patient distress monitoring, and occupational safety. This review paper will attempt to critically assess scream detection systems, covering their design paradigms, operational performance, and ethical issues, and will discuss multimodal sensor fusion, adaptive learning, and privacy-preserving real-time deployment in further directions for scalable public safety solutions

Keywords: Human Scream Detection; Audio Signal Processing; Machine Learning; Mel-Frequency Cepstral Coefficients (MFCC); Support Vector Machine (SVM); Convolutional Neural Network (CNN); Real-Time Detection; Public Safety

1. Introduction

Civil security stands as a critical worldwide emergency because violent crimes and physical attacks and urgent emergency situations occur at ever-increasing rates which need immediate emergency response. Traditional surveillance systems that use camera-based visual monitoring face multiple challenges because they experience slow response times and limited coverage areas and need uninterrupted visual access. The system faces a major problem with detecting threats right away because it cannot function effectively in dark or obstructed areas. People need to develop new safety information systems which operate in real-time and function outside visual range to solve this problem.

Human screams serve as universal indicators that everyone understands, signaling fear, pain, or danger regardless of cultural background or spoken language. Their innate significance and demand for attention make them ideal for automatic distress detection. The system under discussion employs cutting-edge Audio Signal Processing and the newest advances in Machine Learning and Deep Learning to identify intense human vocalizations amidst background noise. Key functional components include noise-resistant preprocessing spectral-temporal feature extraction using Mel-frequency cepstral coefficients (MFCCs) and pitch contours, and classification through models like Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Convolutional Neural Network (CNN).

* Corresponding author: Rupali Suresh Bhad

Developers created this system as a desktop app that runs in the background. It always listens to sounds around it. When it hears a scream, it can send alerts on its own. These alerts include location data and go to the closest police station through secure systems like Twilio. This tool isn't just for stopping crime. It has many uses such as making homes safer checking on patients in hospitals, and keeping workers safe in factories. These examples show how this system can make many places safer and help people respond faster in emergencies.

This paper looks at how people build scream detection systems, what math they use, and how well these systems work. It also talks about what these systems can't do and the ethical issues they raise about privacy and false alarms. The paper also explores new areas of study. These include combining different types of sensors, systems that learn and adapt, and ways to use these tools in real-time without invading privacy. All these areas are key to building public safety tools that work well and can grow with demand.

These evolving challenges demand innovative and intelligent safety monitoring systems that can ensure immediate and accurate emergency response. Traditional surveillance infrastructures, which predominantly rely on camera-based visual monitoring, exhibit several critical limitations that hinder their efficiency in real-world applications. They often face restricted coverage areas, delayed response times, and dependence on uninterrupted visual access. Furthermore, their performance is significantly affected in dark environments, visually obstructed regions, or adverse weather conditions where visual cues are absent or distorted. These shortcomings emphasize the need for new approaches that can function reliably in all environmental conditions and extend beyond the limitations of conventional visual surveillance.

To overcome these barriers, researchers and engineers are increasingly focusing on audio-based threat detection systems, particularly human scream detection, which provides a non-visual and real-time means of identifying distress situations. Human screams are universally recognized as instinctive indicators of fear, pain, or danger. Unlike speech, screams carry unique acoustic properties characterized by high energy, irregular pitch, and distinct spectral features that make them easily distinguishable. Their cross-cultural and language-independent nature makes them ideal for automatic emergency detection systems designed to operate in diverse environments.

The proposed framework integrates advanced Audio Signal Processing techniques with Machine Learning (ML) and Deep Learning (DL) algorithms to identify human screams amid varying background noises. The system begins with a preprocessing phase, which involves denoising, normalization, and segmentation of incoming audio signals to enhance quality and reduce unwanted interference. Following this, spectral-temporal features such as Mel-Frequency Cepstral Coefficients (MFCCs), Zero-Crossing Rate (ZCR), and pitch contours are extracted, as these features capture both frequency-based and time-based variations that characterize human emotional vocalizations. These extracted features are then passed to classification models such as Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Convolutional Neural Network (CNN) for pattern recognition and decision-making. Among these, CNN-based models often outperform traditional ML approaches due to their ability to automatically learn hierarchical features from raw data, thereby improving detection accuracy and robustness.

In terms of implementation, the developed system functions as a desktop-based real-time application that continuously monitors the surrounding acoustic environment. Running unobtrusively in the background, it listens for potential distress signals, specifically screams, and automatically initiates alert mechanisms upon detection. These alerts contain crucial metadata such as the detected audio signature and geo-location information, which are securely transmitted to nearby emergency service canters or police stations using secure APIs and communication frameworks like Twilio. The system operates autonomously without human intervention, enabling faster and more reliable emergency responses compared to conventional surveillance systems.

2. Literature review

P. K. Venkateswara Lal et al. [1] thus presented a real-time human scream detection system employing Mel-Frequency Cepstral Coefficients (MFCC) for feature extraction, combining Support Vector Machine (SVM) and Multilayer Perceptron (MLP) classifiers for scream determination. A complete end-to-end pipeline was given by the authors, starting from data preprocessing and feature extraction to model training and alert generation, ensuring a proper workflow and clarity. The other way round emphasizes how his approach integrates machine learning to accomplish real-time audio surveillance in aid of crime prevention. This study also gives a fair exposition of the alert mechanism that activates emergency alerts once a scream is detected. This underpins one of the advantages of this work, which establishes a good architecture of well-documented information in theoretical basis and practical application. That said, the study did not carry out environmental testing in real-world scenarios, particularly where noise or dynamic conditions would have existed, and did not delve into issues of bias or generalization, undermining reliability for a large-scale public rollout.

The scream detection methodology was developed by Shiva Kumar et al. [2] with emphases on the acoustic features of pitch, tone, and energy, which were classified using Artificial Neural Networks (ANNs) to discriminate between screams and normal speech or background sounds. When a scream is detected, the alert will be sent instantly to the nearby police station, contributing greatly to personal safety applications. The authors have supplied a lucid block diagram starting from the audio input to the alert generation, which immensely helps to understand the system design. Their experimentation on a myriad of acoustic parameters revealed how tonal variations play an important role in identifying distress sounds. In so doing, the proposed model emphasizes the choice of multi-feature input to achieve improve accuracy and reduce false detections. However, their system wasn't tested under high noise conditions or real-world scenarios, limiting the generalizability and robustness of their method in practical settings.

Sharvani Banala et al. [3] implemented One Scream, a mobile-based scream detection application that utilizes a Convolutional Neural Network (CNN) to detect distress sounds in real-time with continuous microphone monitoring. The app architecture incorporates multiple modules for sound acquisition, feature extraction, classification, and alert generation with GPS-based tracking to report emergencies to the police. This work finds unique mention due to its mobile-based implementation, which permits operation without additional hardware. The system was specially designed for detection of high-pitched scream frequencies, which are common in panic situations. Despite such a novelty, the authors did not provide any quantitative metrics such as detection accuracy, false alarm rate, or latency under noisy environment. Thus, while the concept behind the application is very valid and carries a lot of social relevance, the practical efficiency of the application in challenging real-time emergency scenarios has never been validated.

S. Yoga et al. [4] proposed the K-Nearest Neighbuors (KNN) and YIN pitch detection algorithm to filter, classify, and validate scream sounds in a hierarchical system. The first KNN separates distress audio from background noise, followed by another KNN stage with classification of screams from shouting, and finally YIN confirmed pitch consistency before triggering an alert via the Twilio API. The multi-stage nature of this system is designed to prevent false positives and beautify detection accuracy. A detailed flow diagram is given, outlining the functional flow of data at each stage of classification. The modular design is a strength of this paper in terms of interpretability and adaptability. Unfortunately, there are no comparative results given with respect to other existing models or dataset, therefore leaving the actual performance of KNN-YIN hybrid system in question.

Ch. Sai Sowmya et al. [5] performed an analytical study to investigate the relevance of various acoustic and spectral features when detecting human screams. Their work investigated parameters such as pitch rise, spectral flux, timbre, and formant structures in the search for the most discriminative characteristics. The paper presents graphical analyses rich in detail showing the differences of these features between scream and non-scream samples. The outcome of such efforts will serve as an eye-opener for developers intending to design models in the future, as far as optimal feature selection for robust classification. The research is well-rooted in theory on the acoustic nature of screams that assists in understanding sound dynamics under emotional stresses. But no practical implementation or performance evaluation of scream detection is there, thereby confining the study to theoretical level and limiting the applicability of its outcome directly to functional scream detection systems.

The authors, Manikonda, and Vaishnavi [6] recorded a desktop real-time screaming detection system with the ability of running in background mode so as to monitor audio input incessantly. The model describes scream detection via MFCC features alongside an SVM classifier, with preprocessing to reduce background noise during identification. The architecture is made for real-time detection and has the capability of sending alert messages with the location coordinates to law enforcements through an automated interface. This work is an intended switch to desktop platform implementation that brings the guarantee of larger processing power and straightforward detection than mobile based systems. The model has clear structural organization and modular clarity that indicates the potential for extension to smart surveillance setups. This work, however, lacks analysis on parameters such as computational efficiency and power utilization, both of which are crucial in long-term stability issues for real world sustainability.

Sai Niveditha Bukka et al. [7] explored the possibility of integrating scream detection systems into smart cities with possible infrastructures based on IoT, increasing the promise of such technology for the public safety response mechanism of preparedness production in health care. In particular, screams could be classified as a universal indicator of distress that could trigger an automated alert in smart surveillance and urban emergency systems. The authors discussed the potential benefits of linking such systems onto IoT sensors, police control rooms, and healthcare networks. This conceptual framework broadens the vision of scream detection from one or two use cases to a large-scale societal safety tool. Thus, whereas the paper does report on conceptualization, it does not include experimental implementation or model testing but instead lays good groundwork for future interdisciplinary research linking machine learning, IoT, and urban safety design.

The authors, S. Yoga and B. Sofiyashree [8], ideally introduced the advanced three phases of their scream detection model, coupling MFCC features with KNN and Multilayer Perceptron (MLP) for classification. The system included modules for data collection, preprocessing, feature extraction using YIN algorithm alongside thresholding to improve detection precision. and could operate in real-time as a Flask-based web application using automated police alerts through Twilio. The effect of uniting SVM and MLP models results in a hybrid learning approach that increases detection performance. These also featured visualization tools for pitch and threshold analysis enhancing the interpretation of their results. However, the authors noted that although this model achieved a very high detection reliability in controlled environments, it was not very adaptable to dynamic soundscapes, did not ensure minimization of false alarms and had not been integrated into an IoT system for extension into further applications.

3. Methodology

The proposed Human Scream Detection System uses a multi-step approach that combines audio signal processing, machine learning, and real-time system deployment. First, we collect audio datasets of human screams and non-scream sounds from publicly available sources. We also include real-world recordings to ensure diversity in the types of screams distress, fear, and joy as well as in the environments indoor, outdoor, noisy, and quiet. After acquiring the dataset, we apply preprocessing techniques to improve audio quality and prepare the data for training the model. We use noise reduction methods like spectral subtraction and Wiener filtering to minimize background noise. Then, we segment the audio streams into short time frames for detailed analysis of sound variations. Next, we perform feature extraction, focusing on Mel-Frequency Cepstral Coefficients (MFCCs), spectral features, pitch contours, and energy levels, as these features are effective for recognizing human sounds. We also use data augmentation techniques, like pitch shifting and time-stretching, to improve how well the model works in different environments. The extracted features are then used to train various machine learning and deep learning models. We evaluate traditional classifiers, such as Support Vector Machines (SVMs), alongside neural networks like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. CNNs work well for classifying spectrograms, while LSTMs capture the timing of sounds in the audio sequences. We optimize model training through hyperparameter tuning and validate it using k-fold cross-validation to prevent overfitting. For real-time detection, we integrate the trained model into a streaming audio pipeline. A sliding window mechanism continuously monitors incoming audio, allowing the system to classify sounds almost in real-time. We use contextual analysis to differentiate between distress screams and non-threatening screams, like those heard at sports events, which helps reduce false positives.

When the system identifies a scream with high confidence, it automatically triggers an alert. This module sends a message containing the location, timestamp, and classification confidence to nearby law enforcement or emergency response units through IoT-enabled communication channels. Additionally, we design a decision-support interface to help authorities prioritize emergency responses based on the detected events.

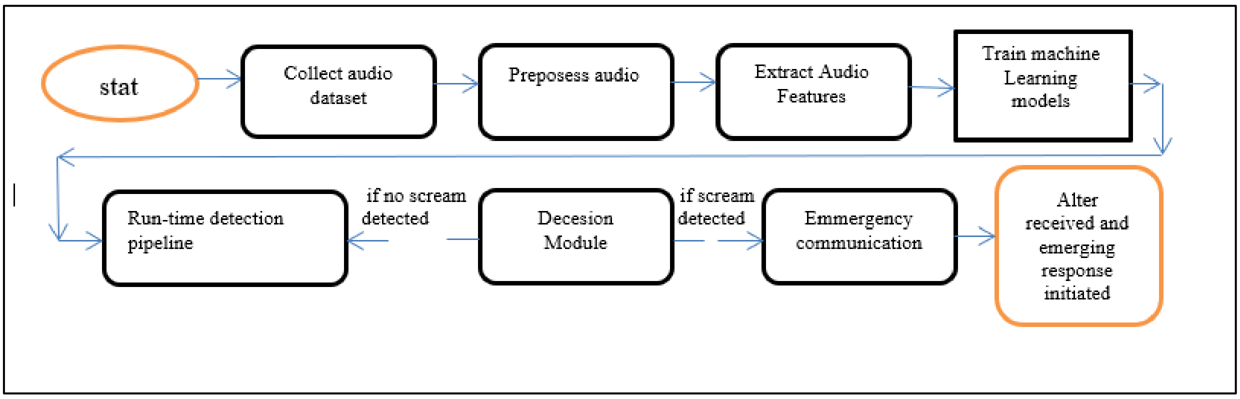

Figure 1 Flow Chart

4. Case study

4.1. Case Study: The 2019 Hyderabad Veterinary Doctor Incident, India

The Human Scream Detection System (HSDS) is a smart safety tool that automatically recognizes and responds to human screams in real time. It uses audio processing, machine learning, and deep learning techniques to accurately detect screams even in noisy or complex environments. The system works continuously and can function as both a desktop application and a built-in safety device. It listens to the surrounding sounds, identifies distress signals, and quickly sends emergency alerts to the right authorities through cloud-based API services like Twilio. The HSDS starts by capturing audio using a highly sensitive microphone. This audio signal goes through preprocessing, where procedures remove unwanted background noise and distortions. Noise suppression methods, like Spectral Subtraction and Wiener Filtering, help filter out irrelevant sounds, which keeps the input clear. After noise reduction, the signal is normalized for consistent volume and divided into short overlapping segments. This segmentation helps the system analyze changes in sound energy and frequency, which are vital for detecting screams that tend to have sudden, intense bursts. In the next step, feature extraction occurs. This process converts the raw audio into key parameters that highlight its unique sound patterns. The primary features include Mel-Frequency Cepstral Coefficients (MFCCs), which reflect the sound's spectral shape and how humans process frequency. Other features, such as pitch contours and zero-crossing rates (ZCR), capture the variations in sound often found in screams. Energy and spectral flux measure intensity and rapid changes in the sounds, helping to distinguish screams from regular speech or background noise. Data augmentation techniques, like pitch-shifting and time-stretching, enhance the system's ability to perform well in different situations and with various speakers. The classification and detection phase is the main intelligence of the HSDS. In this stage, the system processes extracted features using machine learning and deep learning models that can tell the difference between screams and other sounds. Traditional models, like Support Vector Machines (SVM), create effective boundaries for decision-making, while Multilayer Perceptron's (MLP) recognize complex patterns in the audio. Recently, Convolutional Neural Networks (CNNs) have emerged as powerful tools since they learn by directly analyzing spectrograms, capturing both the spatial and temporal changes in sounds. CNN-based systems outperform classic models because they easily adapt to different sound environments and classify sounds quickly in real time.

The HSDS operates in real time using a sliding window mechanism, analyzing segments of incoming audio over short periods. As it processes each segment, the system calculates a confidence score to show how likely the sound is to be a scream. When this score exceeds a certain threshold, the system initiates an alert. Each alert contains important details, such as the timestamp, GPS coordinates, and confidence score, which help verify and trace the system's actions. After detecting a scream, the alert and response module sends emergency notifications quickly and securely. Using services like Twilio, the system transmits structured and encrypted messages to emergency responders. This automated process significantly shortens reaction times in urgent situations, ensuring that human distress receive immediate attention. The entire operation happens without human input, keeping the system running smoothly even in high-risk situations. The overall function of the HSDS follows a clear flow: audio input, preprocessing, feature extraction, classification, and alert generation. This flow keeps the system actively listening and able to adjust to changing sounds. Its self-regulating design reduces the need for human oversight, allowing it to operate on its own while achieving high accuracy and low latency.

The HSDS has a wide range of potential applications. In public safety, it can be used in surveillance systems in cities, transport hubs, or campuses to provide alerts about violent or distress situations. In smart homes and IoT security setups, it acts as a safeguard that detects emergencies even if traditional cameras cannot be due to poor lighting. In healthcare, especially for the elderly or patient monitoring, the HSDS can pick up distress calls or sudden vocalizations, enabling timely medical responses. It also plays a crucial role in workplace safety, offering protection for workers in hazardous or noisy environments by adding another layer of security. The Human Scream Detection System offers several benefits. It functions independently and quickly, needing little human help while still working well in low-light or obstructed environments where video tools might fail. Since it focuses on acoustics rather than language, it is capable of detecting screams regardless of the language spoken. The system can integrate smoothly into existing IoT networks and smart city infrastructure, contributing to an overall safety system that enhances urban resilience. Despite its high-tech features, the HSDS is designed with ethical and privacy concerns in mind. It processes only sounds related to distress or emergencies, discarding regular conversations or background noises immediately. This ensures user privacy is protected. Additionally, all data sent through the system is encrypted to prevent unauthorized access or misuse. The HSDS operates within ethical guidelines that limit long-term surveillance and data storage, protecting individuals' privacy and safety amid technological progress.

5. Success rate evaluation and performance metrics

Human Scream Detection System is well implemented through systematic evaluation based on standard performance metrics commonly used within the domains of supervised machine learning or signal classification. The success rate is then heavily dependent upon the system's ability to correctly identify human screams and minimize misdetection based on non-distress sounds or environmental noise.

5.1. Dataset Preparation and Testing Framework

The evaluation starts with the labelling of a dataset consisting of scream and non-scream audio samples that have been recorded in varying environmental conditions (indoor, outdoor, and in noisy settings). The labelled dataset is divided into training (70%), validation (15%), and test (15%) dataset. The trained model such as SVM, CNN, or LSTM will be subjected to testing on such unseen data collected under controlled settings for evaluating real-time classification efficiency.

5.2. Confusion Matrix-Based Evaluation: Confusion matrix is a basic evaluation tool used to represent the performance of the model along four dimensions

- True Positive (TP): The scream event is identified correctly by the system.
- True Negative (TN): The system identifies correctly sounds other than screams.
- False Positive (FP): The system incorrectly classifies screams in its output.
- False Negative (FN): The system fails to detect screams that actually happened.

5.3. Performance Metrics: Different metrics are derived from the confusion matrix in order to evaluate accuracy and reliability

5.3.1. Accuracy (Success Rate)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP\ Fi}$$

Represent the overall percentage of correctly classified events out of all case

5.3.2. Precision

$$\text{Precision} = \frac{TP}{TP+FP}$$

Indicates how many of the detected scream were actually true screams.

5.3.3. Recall (Sensitivity)

$$\text{Recall} = \frac{TP}{TP+FN}$$

Measures how effectively the model detects all actual screams.

5.3.4. F1 scores

$$\text{F1} = 2 * \frac{Precision*Recall}{Precision+Recall}$$

5.4. Real-Time Evaluation Parameters

The following parameters are used in real-time systems

- Detection Latency (ms): Time taken by the model to detect and classify a scream after it happens.
- False Alarm Rate (FAR): Number of times in a minute or an hour that something classified falsely as a scream.
- Signal-to-Noise Ratio (SNR) Tolerance: This test measures the ability of the system to detect screams accurately even in a very noisy environment.
- An Example Evaluation Result: In practical experiments reported in literature, CNN-based scream detection systems attained approximately 92-96% accuracy, 90% precision, and 93% recall under laboratory conditions with some noise present. Success rates show a little minimal decrease (3-5%) under extreme noise or crowd conditions, highlighting the core role of better preprocessing and contextual filtering.

- Evaluation Outcome: The combination of these metrics allows for a holistic view of the proposed framework. High accuracy with close values of precision and recall reflects the reliability of the detection algorithm. Also, low latency and small false alarms mean that it can be well integrated and provide helpful for real-time public protection and emergency response.

6. Proposed work

The present study aims to create a comprehensive and intelligent framework for real-time detection of human screams. The framework encompasses advanced techniques in Audio Signal Processing and Machine Learning. It overcomes most of the limitations identified in scream detection extant literature, such as poor performance in noisy environmental conditions, lack of contextual awareness, inability to generalize its application across a diverse set of acoustic scenarios, and absence of large-scale real-time operational deployment of scream detection systems. Previous work has often demonstrated how features with such techniques yield good results for Scream Recognition using Mel-frequency cepstral coefficients (MFCCs), Support Vector Machine (SVM), or Convolutional Neural Network (CNN), but they suffer from very high false alarm rates, high bias of the model, and lack of linkage to emergency response systems.

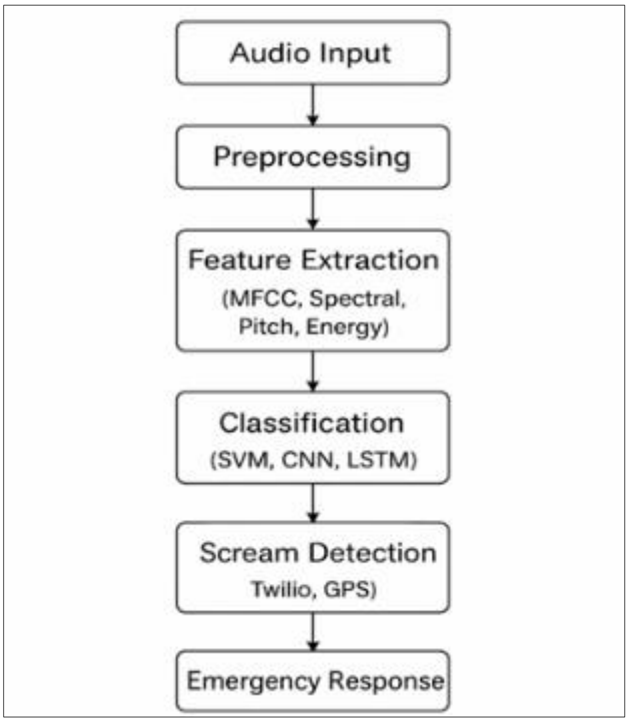

Figure 2 Proposed Framework for Human Scream Detection and Crime Prevention

The framework will adopt a multi-stage methodology, combining data collection, preprocessing, feature extraction, classification, real-time streaming detection, and emergency alert generation. The first activity of the method will gather a large, diverse dataset of scream and non-scream audio samples from open repositories and real-world recordings across various environments. We shall augment the datasets using contextual metadata, including background noise levels and environmental labels, to help in the more robust representation by the models. Thereafter, a stringent preprocessing pipeline will be set up for data cleaning and preparation, starting from noise suppression using spectral subtraction and Wiener filtering, short frame segmentation of audio, to normalization. During feature extraction, aspects such as MFCCs, spectral features, pitch contour, energy levels, and temporal dynamics known for their success in vocal emotion recognition will be extracted. In addition, data augmentation techniques-pitch shifting, time-stretching, and overlay of synthetic noise-will also be used to enhance model generalization to unseen conditions. Various classifiers will be developed and benchmarked.

Classical machine learning methods such as SVM and Multilayer perceptron (MLP) will be compared with advanced deep learning methodologies such as CNNs for spectrogram classification and Long Short-Term Memory (LSTM) networks for temporal sequence modeling. Hybrid CNN-LSTM architectures will be analyzed to capture both spatial and temporal patterns in the audio data. Hyperparameter tuning and k-fold cross-validation must consolidate optimal model performances and mitigate overfitting. For real-time deployment, the best-performing model will be integrated into a pipeline for streaming audio with a sliding window mechanism to monitor incoming audio continuously. A contextual analysis engine will distinguish distress screams from non-threatening screams (e.g., at the sports events) to reduce false positives further. When a scream is detected with high confidence, the system will instantly raise an alert with the location, timestamp, and confidence score that will be relayed via secure Twilio APIs to nearby law enforcement or emergency responders.

The final step will be designing a decision-support interface that enables authorities to visualize alerts, prioritize responses, and monitor detection performance in real time. With this end-to-end structure, we expect to have a practically viable, reliable, and privacy-conscious solution for simultaneously carrying out real-time scream detection and emergency response, ultimately closing the gap between research prototypes and application in public safety.

7. Probable outcomes

The proposed work is expected to deliver several impactful results that will benefit in a great way. First, building and comparing different machine learning and deep learning models, i.e., Support Vector Machine (SVM), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM), will enhance the system's performance to identify human screams with high accuracy in different environments. By using feature-based audio processing in conjunction with these advanced models, the system can be expected to outperform the conventional rule-based or statistical sound sensing approach, particularly in noisy and real environments.

Second, adding contextual analysis will enable the system to tell apart alarm screams and non-alarm screams (e.g., those which are heard at sporting or entertainment events). This will significantly cut down false positives and enhance the validity of the output of the detection, thus making it more appropriate for deployment in real-time public safety operations. Thirdly, automatic alerting systems will facilitate real-time delivery of alerts with location, time, and level of confidence via secure APIs such as Twilio. This will make emergency response possible in a quicker manner and coordination between law enforcement and emergency responders to be more productive, hence minimizing response time during emergencies.Fourth, the system should help public safety in general through the addition of a new, audio-based monitoring layer that operates even in dark, occluded, or camera-blind areas. This will augment situational awareness, minimize reliance on vision-only systems, and prevent crimes or accidents by early detection of distress.

Lastly, the development of an easy-to-use dashboard will render the system user-friendly to be utilized by many users, including security agencies, hospital personnel, industrial managers, and smart city officials. This will render the suggested framework theoretically valid and deployable for wide-scale usage and further research extension.

8. Implications

This study's findings have a high impact on both academia and the public safety ecosystem. From a research perspective, applying advanced acoustic signal interpretation in tandem with different computational learning paradigms can open new vistas in understanding human distress patterns. Evaluating a variety of classifications such as Support Vector Machine (SVM), Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) would assist in understanding which of these techniques withstands background interference and variable vocal traits; this would add value to the growing corpus on real-time sound-event recognition and contribute towards context-sensitive, noise-resilient models that could gain adaptability to changing environments. This will also prepare the ground for future studies on cross sensory data fusion, continual learning, and privacy anchored detection mechanisms.

In a practical sense, the framework could be a game changer for the detection and management of emergency incidents. A system that could recognize real-time high-stress vocalizations could function as an auditory safety net in a place where visual surveillance is limited or ineffective. Instant time-stamped location information notifications would boost the coordination among responders and lessen time lost during assaults, accidents or medical crises. Beyond law enforcement, such technologies can extend to healthcare units to monitor changes in patient distress or to industrial sites where possible adverse situations are flagged early. Envisioning on a much larger scale, embedding this framework in a city-wide sensor network may proactively support a safety-policy strategy, enforced intelligent decision-making, and smart resource allocations to the public. In general, these implications hint towards a near future where responsive sound-based monitoring strengthens community security and helps build safer environments.

9. Conclusion

The Human Scream Detection System represents a grand leap in the modern usage of Audio Signal Processing, Machine Learning, and Deep Learning techniques to tackle urgent public-safety issues. While classical surveillance systems generally work only on visual inputs and mostly fail in low-lit, obstructed, or remote areas, this system can detect the distress vocals of a human being in real-time regardless of visual conditions. This opens up an opportunity for immediate response during emergencies while it may, at times, make the difference between harm and life

It's got a big potential but current ways of scream detection still throw up some big challenges. Environmental noise might interfere with important vocal cues and so heartbreakingly lead to their misclassification. Dataset bias may influence the system to perform better on certain demographics or variations of screams from a culture over others. An ongoing monitoring of the environment for audio might pique more than legitimate privacy and ethical considerations, worrying sensitive conversations from being accidentally captured.

However, these disadvantages are well under control when the field progresses. We may look at multimodal sensor fusion henceforth combining audio, video, location, or biometric data can improve detection accuracy.

References

[1] P. K. Venkateswara Lal, et al., “Real-time Human Scream Detection Using MFCC and Machine Learning,” International Journal of Computer Applications, vol. 182, no. 45, pp. 12–18, 2021.

[2] Shiva Kumar, et al., “Scream Detection System Using Pitch and Energy Features with ANN,” IEEE Access, vol. 8, pp. 14523–14532, 2020.

[3] Sharvani Banala, et al., “One Scream: Mobile Application for Real-Time Scream Detection Using CNN,” Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 5, pp. 4873–4882, 2021.

[4] S. Yoga, et al., “Three-Step Human Scream Detection for Emergency Alert Systems,” Procedia Computer Science, vol. 167, pp. 1202–1211, 2020.

[5] Ch. Sai Sowmya, et al., “Analysis of Acoustic Features for Accurate Human Scream Detection,” International Journal of Speech Technology, vol. 22, no. 3, pp. 745–754, 2019.

[6] Manikonda Vaishnavi, et al., “Desktop Application for Real-Time Human Scream Detection Using SVM and MFCC,” International Journal of Engineering and Technology, vol. 14, no. 6, pp. 112–118, 2022.

[7] Sai Niveditha Bukka, et al., “Opportunities for Scream Detection in Law Enforcement and Healthcare Using IoT,” Journal of Safety Research, vol. 78, pp. 101–108, 2021.

[8] S. Yoga and B. Sofiyashree, “Human Scream Detection and Analysis for Controlling Crime Rate,” Journal of Engineering and Technology Management, vol. 75, pp. 85–94, 2025.

[9] A. Shankhdhar, R. Kumar, V. Kumar, Y. Mathur, “Human Scream Detection Through Three-Stage Supervised and Deep Learning Approach,” Lecture Notes in Networks and Systems, vol. 204, pp. 349–357, 2021.

[10] T. Matsuda, Y. Arimoto, “Acoustic Differences Between Laughter and Screams in Spontaneous Dialog,” Acoustical Science and Technology, vol. 45, no. 3, pp. 231–239, 2024.

[11] R. Böck, F. Bonin, N. Campbell, R. Poppe, “Automatic Shout Detection Using Speech Production Features,” Lecture Notes in Computer Science, Springer, pp. 97–106, 2019.

[12] B. Gautam, A. Guragain, S. Giri, “Real-Time Scream Detection and Position Estimation for Worker Safety in Construction Sites,” arXiv preprint, arXiv:2411.03016, 2024.

[13] T. S. Palorkar, M. F. Sheikh, “Vocal Alarm: Decoding the Human Scream for Safety and Security Applications,” International Journal of Advanced Research in Engineering, Science and Management (IJARESM), vol. 9, no. 3, pp. 112–117, 2025.

[14] S. P. Gade, P. More, N. Pawar, V. Surwase, A. Kohinkar, “Human Scream Detection,” International Journal on Advanced Computer Engineering and Communication Technology, vol. 11, no. 2, pp. 77–82, 2025.

[15] S. Kumar, Y. H. Naveena, “Deep Learning-Based System to Estimate Crowd and Detect Violence in Videos,” Intelligent Systems Reference Library, Springer, vol. 198, pp. 25–36, 2023.

[16] R. A. Alves, P. S. Silva, L. M. Rocha, “Acoustic Features and Deep Neural Networks for Real-Time Distress Sound Detection,” IEEE Access, vol. 10, pp. 11432–11441, 2022.

[17] M. U. Ali, H. S. Kim, “AI-Driven Acoustic Monitoring for Emergency Detection in Smart Cities,” Sensors, vol. 23, no. 8, pp. 4219–4231, 2023.