International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

AI Enabled Human Scream Detection And Safety Alert System

Rupali Suresh Bhad[1], Dr. Harsha R. Vyawahare[2], Dr. A. A. Khodaskar[3]

[1]Research Scholar, MTech Computer Science and Engineering, SIPNA College of Engineering and Technology, Amravati.
[2,3]Professor, Computer Science and Engineering, SIPNA College of Engineering and Technology, Amravati.

Abstract
Human safety and crime prevention require intelligent systems capable of detecting emergency situations in real time. Traditional surveillance systems mainly depend on visual monitoring, which may be ineffective in low-light conditions or obstructed environments. Human screams are universal indicators of fear, danger, and distress, making acoustic monitoring an effective approach for emergency detection. This paper presents an AI Based Human Scream Detection System for Crime Prevention that utilizes deep learning techniques to identify distress signals from environmental audio. The proposed system employs Mel Spectrogram-based feature extraction and a ResNet34 Convolutional Neural Network (CNN) to classify audio signals into scream and non-scream categories. The framework supports both live microphone monitoring and audio file analysis through a Flask-based web application. Audio signals undergo preprocessing, normalization, and spectrogram transformation before classification by the trained model. Detection results are displayed through an interactive dashboard that provides real-time monitoring and alert generation. Experimental results demonstrate a classification accuracy of approximately 87.7% with low inference latency, making the system suitable for near real-time applications. The proposed framework offers a practical and scalable solution for deployment in smart surveillance systems, educational institutions, healthcare facilities, workplaces, and smart city environments to improve public safety and emergency response.

Keywords- Human Scream Detection, Crime Prevention, Deep Learning, ResNet34, Mel Spectrogram, Acoustic Event Recognition, Smart Surveillance.

1. Introduction

The rapid growth of urban populations, increasing crime rates, workplace accidents, and public safety concerns have created a strong demand for intelligent monitoring systems capable of detecting emergency situations in real time. Traditional surveillance systems mainly rely on cameras and human supervision, which often become ineffective in low-light conditions, visually obstructed areas, or locations beyond camera coverage. As a result, there is a growing need for alternative monitoring approaches that can identify distress situations without depending solely on visual information. Human screams are universal indicators of fear, danger, pain, and emergency. Unlike visual signals, screams can be detected even when the source is not directly visible, making them valuable acoustic cues for

emergency identification. Automatic recognition of distress-related sounds can significantly improve public safety by enabling faster intervention and response during critical situations. Such systems can assist security personnel, healthcare providers, emergency responders, and law enforcement agencies in identifying incidents that might otherwise remain unnoticed.

Recent advancements in artificial intelligence, machine learning, and deep learning have transformed the field of audio event recognition. Modern acoustic monitoring systems can analyze environmental sounds and distinguish specific events from complex background noise. Earlier approaches relied on handcrafted audio features and conventional machine learning algorithms, which often struggled to maintain accuracy in dynamic real-world environments. The emergence of deep learning, particularly Convolutional Neural Networks (CNNs), has improved audio classification by automatically learning discriminative features from spectrogram representations and reducing dependence on manual feature engineering. This paper presents an AI-Based Human Scream Detection System for Crime Prevention, a deep learning-powered framework designed to identify distress signals from environmental audio. The proposed system utilizes Mel Spectrogram representations and a ResNet34 CNN architecture to classify audio signals into scream and non-scream categories. The framework supports both live microphone monitoring and offline audio analysis, providing flexibility for real-world deployment. By combining deep learning, acoustic signal processing, and real-time monitoring, the system offers an effective solution for enhancing public safety, emergency response, and crime prevention applications.

2. Literature Review

Human scream detection has gained significant attention in the fields of acoustic event recognition, intelligent surveillance, and public safety. Early research primarily focused on using handcrafted acoustic features such as Mel-Frequency Cepstral Coefficients (MFCC), pitch, energy, and spectral characteristics combined with machine learning algorithms for scream classification. P. K. Venkateswara Lal et al. developed a real-time scream detection system using MFCC features and machine learning techniques, while S. Kumar et al. utilized pitch and energy-based features with Artificial Neural Networks (ANN) for emergency detection [1], [2]. Although these approaches demonstrated promising results, their performance was often affected by environmental noise and varying recording conditions. With the advancement of deep learning, researchers began employing Convolutional Neural Networks (CNNs) to automatically learn acoustic patterns from audio signals. S. Banala et al. proposed a mobile-based scream detection application using CNNs for real-time monitoring, whereas Shankhdhar et al. introduced a hybrid supervised and deep learning framework that improved classification accuracy [3], [9]. Similarly, Alves et al. demonstrated that deep neural networks outperform traditional machine learning models for distress sound recognition in noisy environments [16].

Several studies have also explored the integration of scream detection systems with safety and surveillance applications. Bukka et al. highlighted the role of IoT-enabled acoustic monitoring in healthcare and law enforcement, while Gautam et al. applied scream detection techniques for worker safety in construction environments [7], [12]. Furthermore, Ali and Kim proposed AI-driven acoustic monitoring systems for smart city infrastructures to support automated emergency response mechanisms [17].

Recent advancements in environmental sound classification have further strengthened audio-based surveillance systems. Researchers such as Piczak, Hershey et al., and Salamon et al. demonstrated the effectiveness of CNN-based architectures for recognizing complex environmental sounds using spectrogram representations [18], [19], [20]. These studies established the foundation for modern deep learning-based acoustic monitoring frameworks. Despite significant progress, existing systems often suffer from limitations such as false alarms, poor adaptability to diverse sound environments, limited scalability, and lack of integrated monitoring interfaces. To overcome these challenges, the proposed AI Based Human Scream Detection System for Crime Prevention utilizes Mel Spectrogram feature extraction and a ResNet34 deep learning architecture integrated with a Flask-based web dashboard. This approach provides improved classification accuracy, real-time monitoring capability, and enhanced usability for practical deployment in intelligent safety and crime prevention applications.

3. Methodology

The proposed **AI-Based Human Scream Detection System for Crime Prevention** follows a systematic methodology to identify distress signals from environmental audio using deep learning techniques. The framework consists of three major phases: audio processing, deep learning-based classification, and alert generation. This structured approach enables accurate scream detection and real-time monitoring for safety applications.

3.1 Audio Acquisition and Preprocessing

The first stage involves collecting audio through live microphone monitoring or uploaded audio files. The acquired audio signals are preprocessed to improve quality and consistency. Normalization is applied to standardize signal amplitude, while audio samples are padded or trimmed to a fixed duration suitable for model input. Basic noise handling techniques are also used to reduce environmental interference. The processed audio is then converted into Mel Spectrogram representations, which capture important spectral and temporal characteristics of sound signals and serve as inputs to the classification model.

3.2 Deep Learning-Based Classification

In this phase, the generated Mel Spectrogram images are analyzed using a ResNet34 Convolutional Neural Network (CNN). The model automatically extracts meaningful acoustic features and classifies audio samples into scream and non-scream categories. The residual learning architecture of ResNet34 improves feature extraction efficiency and classification accuracy. During inference, the trained model generates prediction probabilities that indicate the likelihood of a distress signal being present in the audio sample.

3.3 Alert Generation and Monitoring

The final phase evaluates the classification output using confidence-based decision logic. If the predicted scream probability exceeds a predefined threshold, the system identifies the event as a distress signal. An alert is immediately generated and displayed through the monitoring dashboard. The detected audio clip, prediction result, and timestamp information are stored for future reference. The Flask-based dashboard

provides real-time visualization of detection statistics and activity logs, enabling effective monitoring and rapid response during emergency situations.

The integration of audio processing, deep learning classification, and real-time monitoring creates an intelligent framework capable of supporting crime prevention and public safety applications.

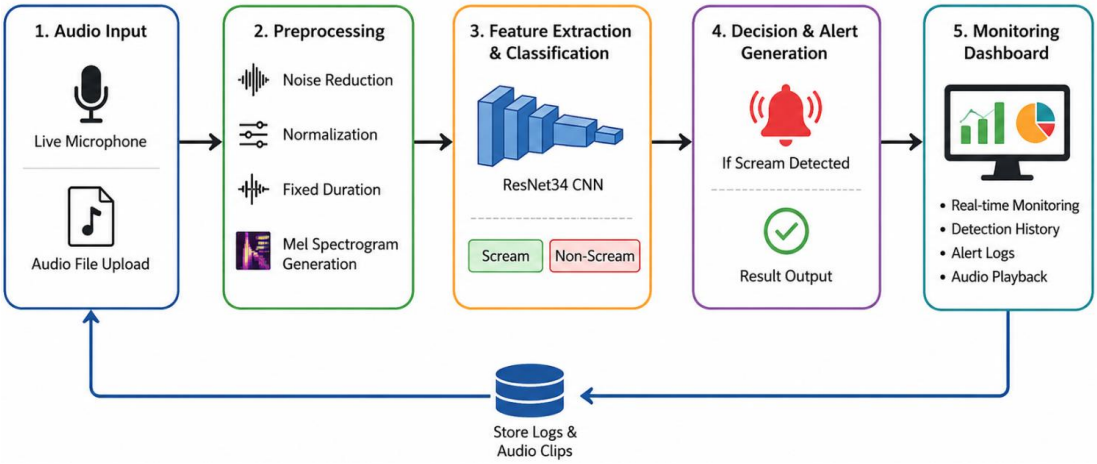

Figure 3.1: AI-based crime prevention system flowchart

4. Proposed System

The proposed AI Based Human Scream Detection System for Crime Prevention is designed to automatically identify distress signals from environmental audio and generate alerts in real time. The system integrates audio signal processing, deep learning-based classification, and web-based monitoring to provide an intelligent framework for enhancing public safety and emergency response. The overall architecture consists of audio acquisition, preprocessing, deep learning classification, alert generation, and dashboard visualization modules.

4.1 System Architecture

The proposed system follows a structured workflow that begins with audio acquisition and ends with real-time monitoring. Audio signals are collected either through live microphone monitoring or uploaded audio files. The captured audio is processed and transformed into Mel Spectrogram representations, which are then analyzed using a ResNet34 Convolutional Neural Network (CNN). Based on the classification results, the system generates alerts and displays the output through a web-based dashboard. The architecture is designed to ensure efficient processing, low latency, and scalability. Each module operates independently while maintaining seamless communication with other components. This modular design improves maintainability and supports future enhancements such as cloud deployment and IoT integration.

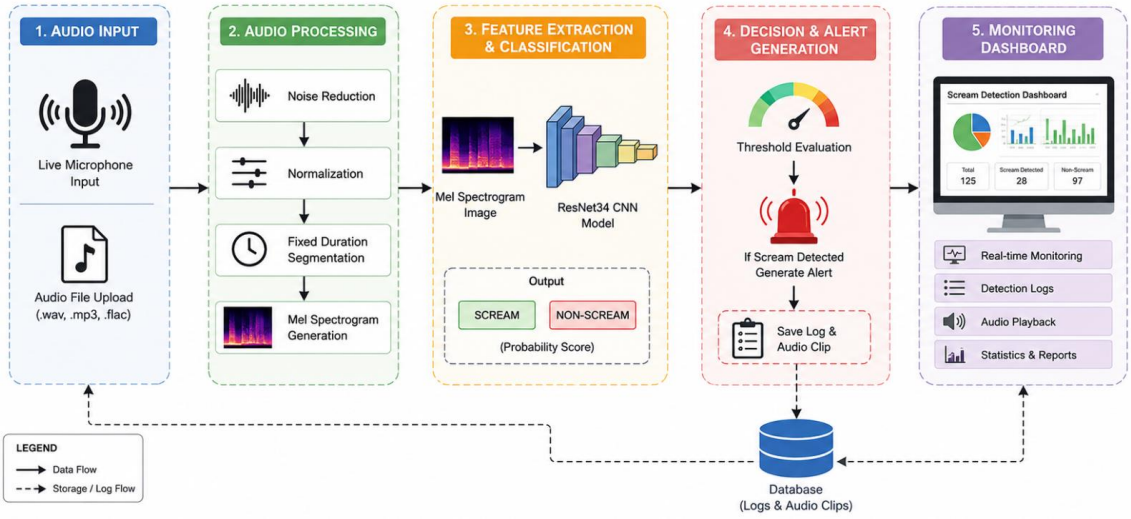

Figure 4.1: Proposed System Architecture of AI Based Human Scream Detection System for Crime Prevention

4.2 Audio Processing and Feature Extraction

The audio processing module is responsible for preparing environmental audio signals for classification. The system accepts audio input from live microphone streams and uploaded audio files in multiple formats. After acquisition, preprocessing operations such as normalization, duration standardization, and basic noise handling are performed to improve signal quality.

The processed audio signals are converted into Mel Spectrogram images using the Librosa library. Mel Spectrograms provide a frequency-domain representation of audio signals and effectively capture both temporal and spectral characteristics of distress sounds. These spectrogram images serve as input to the deep learning model and provide richer information compared to traditional handcrafted features such as MFCC alone.

4.3 Deep Learning Classification and Alert Generation

The classification module utilizes a ResNet34 Convolutional Neural Network trained on labeled scream and non-scream audio samples. The model automatically extracts hierarchical acoustic features from Mel Spectrogram images and performs binary classification. The output of the model consists of prediction probabilities indicating the likelihood of a scream event.

Based on predefined confidence thresholds, the system determines whether the detected sound represents a distress signal. When a scream is detected, an alert is generated automatically and displayed on the monitoring dashboard. Simultaneously, the detected audio clip and timestamp information are stored for future analysis. This mechanism ensures timely identification of emergency situations and supports rapid response by monitoring personnel.

4.4 Dashboard Monitoring and Result Visualization

The final module of the proposed system is the web-based monitoring dashboard developed using Flask, HTML, CSS, and JavaScript. The dashboard provides a user-friendly interface for viewing classification results, monitoring live detection activities, and accessing historical records.

Users can upload audio files, initiate live monitoring sessions, review detection logs, and listen to stored audio clips through the dashboard. Real-time visualization of detection statistics improves system usability and enables efficient supervision of monitoring operations. The integration of alert generation and activity logging further enhances the practicality of the proposed framework for crime prevention and intelligent safety monitoring applications.

5. Implementation

5.1 Overview

The implementation of the proposed **AI** Based Human Scream Detection System for Crime Prevention focuses on developing a real-time acoustic monitoring framework capable of identifying distress signals from environmental audio. The system integrates audio processing, deep learning-based classification, alert generation, and web-based visualization into a unified platform. The architecture follows a modular design that ensures scalability, maintainability, and efficient real-time operation. The overall workflow consists of user interaction, audio acquisition, preprocessing, feature extraction, classification, alert generation, and dashboard monitoring.

5.2 User Interface and Audio Acquisition Module

The implementation begins with the user interface, which provides facilities for live microphone monitoring and audio file uploads. Users can upload prerecorded audio files or initiate real-time monitoring through the dashboard. The audio acquisition module captures environmental audio either from a microphone or uploaded files. The system supports multiple audio formats including WAV, MP3, FLAC, OGG, and M4A. The acquired audio signals are then forwarded to the preprocessing stage for further analysis.

5.3 Audio Preprocessing and Feature Extraction

The preprocessing module standardizes audio signals before classification. Operations such as normalization, resampling, noise handling, and duration standardization are applied to improve signal quality and consistency. After preprocessing, the audio signals are converted into Mel Spectrogram representations using the Librosa library. These spectrogram images capture important spectral and temporal characteristics of scream sounds and provide meaningful features for deep learning-based analysis.

5.4 Classification Module Using ResNet34

The core component of the system is the ResNet34 Convolutional Neural Network implemented using the PyTorch framework. The model is trained on approximately 3,537 labeled scream and non-scream audio samples. ResNet34 automatically learns hierarchical acoustic features from Mel Spectrogram images and performs binary classification. During inference, the model generates prediction probabilities for scream and non-scream classes, which are used for decision-making.

5.5 Decision Logic and Alert Generation

The classification results are evaluated using confidence-based threshold logic. If the predicted scream probability exceeds the predefined threshold, the system identifies the audio as a distress event. Upon detection, an alert is generated, the audio clip is stored, timestamp information is recorded, and detection logs are updated. This mechanism enables rapid identification of emergency situations and supports timely response actions.

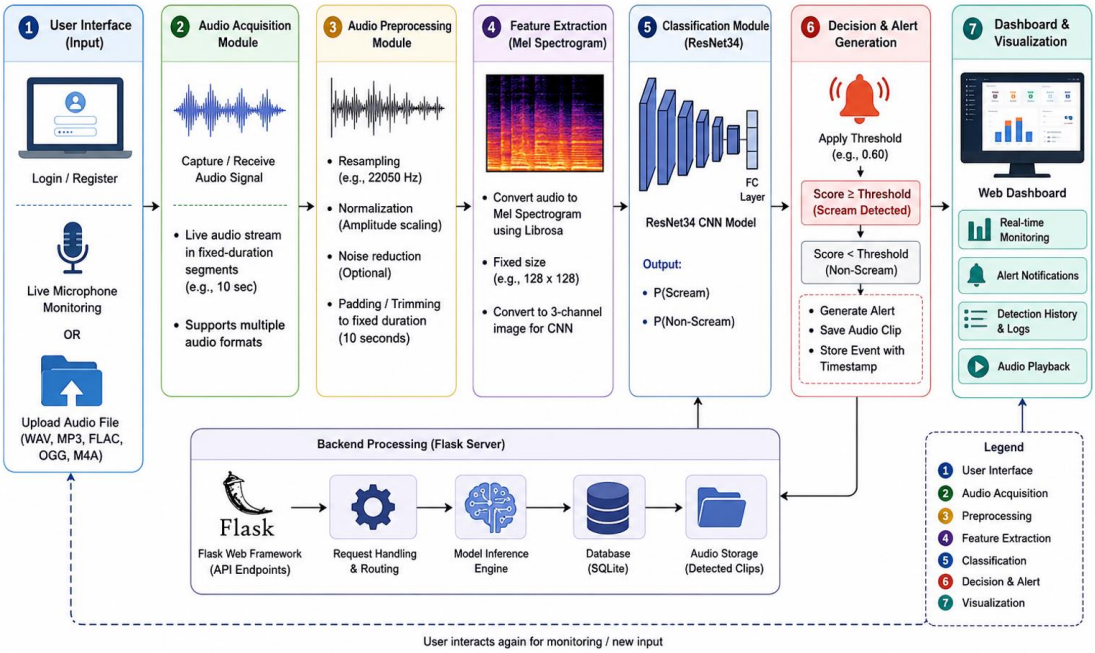

Figure 5.1: Implementation Architecture of AI Based Human Scream Detection System for Crime Prevention

5.6 Backend Processing and Data Management

The backend is developed using the Flask web framework, which manages routing, audio uploads, model inference, and communication between frontend and backend modules. Detection records, timestamps, and event logs are stored in an SQLite database, while detected audio clips are maintained in a dedicated storage repository. This architecture ensures efficient data management and smooth system execution.

5.7 Dashboard Visualization Module

The dashboard serves as the monitoring interface of the system and is developed using HTML, CSS, JavaScript, and Flask templates. It provides real-time monitoring status, alert notifications, detection history, event logs, audio playback functionality, and system statistics. Users can review detection results and historical records through an intuitive interface, making the framework suitable for deployment in surveillance, healthcare, educational, industrial, and smart city environments.

6. Result

6.1 Experimental Setup

The experiments were conducted using Python, PyTorch, Librosa, Flask, and the ResNet34 deep learning model. The dataset consisted of approximately 3,537 labeled audio samples, categorized into scream and non-scream classes. An 80:20 train-test split was used for model development and evaluation.

Before training, all audio samples underwent preprocessing operations such as normalization, duration standardization, and Mel Spectrogram generation. The ResNet34 model was trained using the Adam optimizer with a learning rate of 0.001 and Cross Entropy Loss as the objective function. Early stopping was applied to improve model generalization and prevent overfitting.

6.2 Classification Performance Analysis

The proposed system was evaluated using Accuracy, Precision, Recall, and F1-Score metrics. Experimental results showed a training accuracy of 97–98% and a testing accuracy of 87.7%, demonstrating the effectiveness of the ResNet34 model in distinguishing scream sounds from non-scream audio.

The slight difference between training and testing accuracy is mainly due to environmental noise and acoustic variations present in real-world recordings. Despite these challenges, the model maintained stable performance across diverse sound conditions. The obtained results confirm that the proposed framework provides reliable and robust human scream detection suitable for real-time crime prevention and emergency monitoring applications.

Metric Value		Metric Value	
Training Accuracy	97–98%	Training Accuracy	97–98%
Testing Accuracy	87.7%	Testing Accuracy	87.7%
Precision High		Precision High	
Recall High		Recall High	
F1-Score Balanced		F1-Score Balanced	
Classification Type		Scream / non-scream	

Table 6.1: Performance Evaluation Metrics

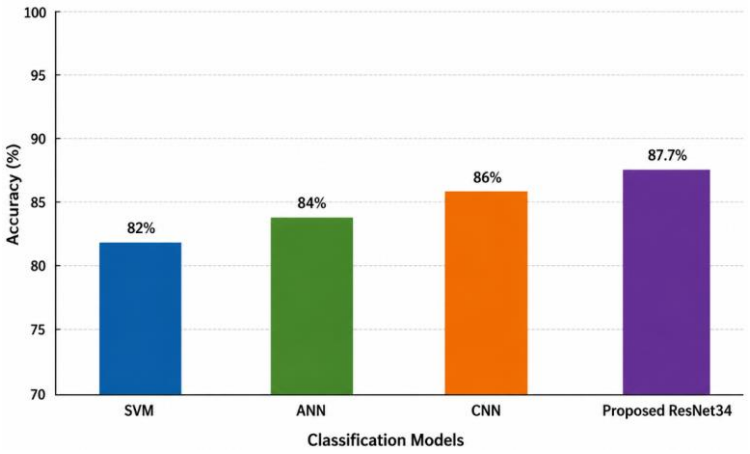

Figure 6.2: Accuracy Comparison of Different Classification Models

The high precision value indicates that the model effectively reduces false alarms by correctly distinguishing scream sounds from other environmental noises. Similarly, strong recall performance demonstrates the system's ability to identify actual distress events successfully.

6.3 Real-Time Performance Evaluation

An important requirement of the proposed system is real-time operation. Therefore, inference latency was evaluated to determine the responsiveness of the detection framework. The time required for audio preprocessing, Mel Spectrogram generation, and deep learning inference was measured during experimentation. The proposed ResNet34 model demonstrated an average inference time of less than 800 milliseconds on CPU-based systems, while significantly lower latency was observed when GPU acceleration was available. This performance confirms that the framework is capable of near real-time scream detection and can generate alerts promptly during emergency situations. The system successfully analyzed live microphone inputs in fixed-duration segments and generated detection results without noticeable delays. This capability is particularly important for surveillance applications where rapid response is critical.

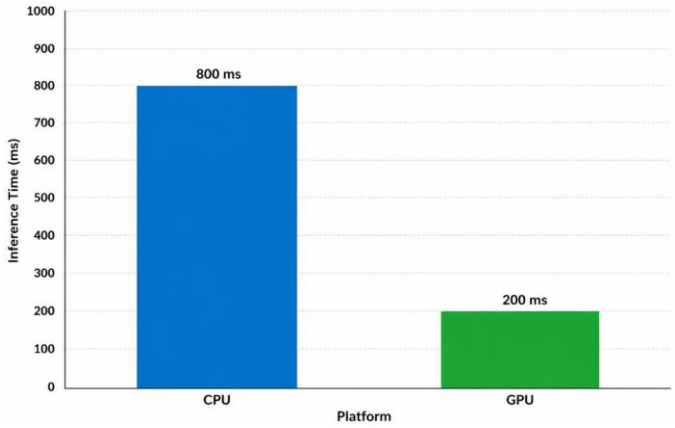

Figure 6.3 : Inference Latency Comparison of CPU and GPU Execution

7. Conclusion

This paper presented an AI Based Human Scream Detection System for Crime Prevention that utilizes deep learning techniques to identify distress signals from environmental audio. The proposed framework integrates audio preprocessing, Mel Spectrogram-based feature extraction, and a ResNet34 Convolutional Neural Network to classify audio signals into scream and non-scream categories with high accuracy. The system supports both live microphone monitoring and audio file analysis, making it suitable for real-time surveillance applications. A Flask-based web dashboard was developed to provide real-time monitoring, alert generation, event logging, and visualization of detection results. Experimental evaluation demonstrated that the proposed system achieved approximately 87.7% testing accuracy with low inference latency, confirming its effectiveness in detecting distress situations under different operating conditions. The developed framework offers a practical and scalable solution for deployment in educational institutions, healthcare facilities, workplaces, public surveillance environments, and smart city infrastructures. By enabling early identification of potential emergency situations, the system can support faster response mechanisms and enhance public safety. Overall, the proposed approach demonstrates the potential of artificial intelligence and acoustic monitoring technologies in developing intelligent crime prevention and emergency detection systems.

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. S. Kumar et al., “Scream detection system using pitch and energy features with ANN,” IEEE Access, vol. 8, pp. 14523–14532, 2020.
3. S. 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. S. 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. M. 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. S. N. 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, and 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 and 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, and R. Poppe, “Automatic shout detection using speech production features,” in Lecture Notes in Computer Science, Springer, pp. 97–106, 2019.

12. B. Gautam, A. Guragain, and 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 and 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, and 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 and 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, and 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 and H. S. Kim, “AI-driven acoustic monitoring for emergency detection in smart cities,” Sensors, vol. 23, no. 8, pp. 4219–4231, 2023.
18. K. J. Piczak, “Environmental sound classification with convolutional neural networks,” in Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP), Boston, MA, USA, 2015, pp. 1–6.
19. S. Hershey, S. Chaudhuri, D. P. W. Ellis, J. F. Gemmeke, A. Jansen, C. Moore, M. Plakal, D. Platt, R. A. Saurous, R. Seybold, M. Slaney, R. Weiss, and K. Wilson, “CNN architectures for large-scale audio classification,” in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017, pp. 131–135.
20. J. Salamon and J. P. Bello, “Deep convolutional neural networks and data augmentation for environmental sound classification,” IEEE Signal Processing Letters, vol. 24, no. 3, pp. 279–283, 2017.
21. Y. Tokozume and T. Harada, “Learning environmental sounds with end-to-end convolutional neural network,” in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017, pp. 2721–2725.
22. K. Chachada and C. J. Kuo, “Environmental sound recognition: A survey,” APSIPA Transactions on Signal and Information Processing, vol. 3, pp. 1–15, 2014.
23. J. F. Gemmeke, D. P. W. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, “Audio Set: An ontology and human-labeled dataset for audio events,” in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017, pp. 776–780.
24. A. Mesaros, T. Heittola, and T. Virtanen, “A multi-device dataset for urban acoustic scene classification,” in Proceedings of the Detection and Classification of Acoustic Scenes and Events (DCASE) Workshop, Munich, Germany, 2017, pp. 9–13.
25. H. Phan, P. Koch, F. Katzberg, M. Maass, R. Mazur, and A. Mertins, “Audio event detection with deep neural networks: A survey,” Artificial Intelligence Review, vol. 54, no. 2, pp. 1071–1120, 2021.