Human Scream Detection and Analysis for Crime Reduction

Mrs. Shruthi TV[1], Aishwarya S Koti[2], Aptha H P[3], Isra Didagur[3]

[1]Shruthi TV , Assoc. Prof, Dept. of ISE, East West Institute of Technology
[2]Aishwarya S Koti, Dept. of ISE, East West Institute of Technology
[3]Aptha H P, Dept. of ISE, East West Institute of Technology
[4]Isra Didagur, Dept. of ISE, East West Institute of Technology

---***---

ABSTRACT - This paper presents an intelligent system that uses Machine Learning to detect crimes in real-time. It analyzes social media feeds and audio sensors to identify distress signals like screams. When a potential crime is recognized, the system automatically sends SMS alerts with location data to the nearest police stations. Field trials showed a 50% reduction in police response time and a 40% increase in successful interventions. This proactive approach, combining real-time analytics and automated alerts, enhances public safety and offers a model for future smart city security frameworks.

Key Words: Scream Detection, Machine Learning, Deep Learning, MFCC Features, Emergency Alert System

1. INTRODUCTION

Ensuring public safety has become an increasingly critical concern in modern society, particularly with the rising need for intelligent surveillance technologies. Traditional security systems often rely heavily on visual monitoring, which demands continuous human supervision and may fail in situations involving poor visibility, camera blind spots, or incidents occurring beyond the camera's field of view. To address these limitations, this project introduces an intelligent audio surveillance system designed to enhance real-time situational awareness and emergency responsiveness.

This innovative system utilizes advanced machine learning algorithms and signal processing techniques to automatically detect and analyze human screams—one of the most immediate indicators of distress or danger. By capturing and processing acoustic signals in real time, the system distinguishes human screams from background noise using unique acoustic signatures such as frequency distribution, amplitude dynamics, and temporal variations. These characteristics enable the system to accurately differentiate between genuine distress calls and non-threatening environmental sounds.

The primary objective of the proposed model is to establish an automated early-warning mechanism that significantly reduces response time during emergencies, thereby potentially preventing crimes and saving lives. The system is designed to operate seamlessly alongside existing surveillance infrastructures, including CCTV networks, emergency alert systems, and law enforcement databases. By integrating audio-based detection into conventional security environments, it provides an additional layer of safety that complements visual monitoring, ensuring a more robust and comprehensive security framework for public and private spaces.

2. LITERATURE SURVEY

2.1 "Identification of Illicit Activities & Scream Detection using Computer Vision & Deep Learning" R.Mathur,T.Chintala,D.Rajeswari (2022)

The study by Mathur et al. (2022) presents an intelligent surveillance system that integrates computer vision with deep learning–based audio analysis to improve public safety. The visual component uses CNN models to detect suspicious or illicit activities such as aggression, theft, or violent movements in CCTV footage, reducing the need for continuous human monitoring. Alongside this, the system incorporates a scream detection module capable of classifying audio into normal sounds, mild distress, and high distress, achieving an accuracy of 95%.

A key strength of the research is its multimodal approach, combining both video and audio cues for more reliable emergency detection. This fusion enables the system to recognize distress situations even when incidents occur outside the camera's view or under low-visibility conditions. The model provides real-time alerts to security personnel or law enforcement, helping reduce response times and enhancing crime prevention. Overall, the framework demonstrates a scalable and effective solution for automated surveillance in public spaces.

2.2 "Detecting Voice Features for Criminal Case" Aie Su Su Ky , K Zin Lin (2021)

The study by Aie Su Su Ky and K Zin Lin (2021) explores how voice features can be used in criminal investigations to detect emotional states and support truth verification. The research highlights that vocal characteristics—such as pitch, tone, rhythm, and

energy—change involuntarily with emotions, making them reliable indicators of a person's psychological state. By analysing these subtle variations in speech, the system aims to distinguish between normal and sad emotional conditions, which can be useful in legal settings where emotional cues may be difficult to interpret.
The proposed model integrates key audio-processing techniques such as Pitch Detection, Mel-Frequency Cepstral Coefficients (MFCCs), and Short-Time Energy (STE) to extract meaningful features from voice samples. These features allow the system to recognize emotional states with improved accuracy, making it a valuable tool for law enforcement, lawyers, and judges during interrogations or courtroom evaluations. Overall, the paper demonstrates how voice-based emotion analysis can provide objective insights in criminal cases, enhancing forensic investigations and contributing to more accurate and fair legal decisions.

2.3 "An Expert System for Identification of Domestic Emergency based on Normal and Abnormal Sound", Swati Shilaskar; Shripad Bhatlawande; Aditya Vaishale; Prapti Duddalwar; Aditya Ingale

The research paper "An Expert System for Identification of Domestic Emergency based on Normal and Abnormal Sound" presents an intelligent audio-based monitoring system designed to enhance safety within home environments. The system continuously listens to everyday household sounds and uses Topological Data Analysis (TDA) to differentiate between normal activities and potentially dangerous anomalies, such as glass breaking, gunshots, screaming, or sudden silence. This ability to detect both loud emergencies and unusual quietness is especially valuable for protecting vulnerable groups like the elderly, children, and people with disabilities.
To achieve accurate sound classification, the researchers evaluated multiple machine learning models including SVM, Random Forest, and Neural Networks. Among these, the K-Nearest Neighbours (KNN) algorithmrealmprovedSuccess achieved the best results, delivering 98% precision in identifying emergency sounds. The high accuracy and low false-alarm rate make the system reliable for real-time alerting, enabling timely intervention by caregivers or emergency responders. Overall, the study demonstrates a powerful approach to domestic safety by integrating advanced audio analysis and machine learning into a non-invasive, always-active smart home emergency detection system.

2.4"Human Scream Detection and Analysis for Crime Reduction", Srimathi A; Jothi S; Sindhiya M P

Crime often goes unreported or is reported too late, making it difficult for law enforcement to respond quickly. To address this issue, this paper proposes a real-time scream detection system that uses machine learning to identify human screams from environmental sounds. The system extracts audio features using MFCCs and classifies them with Support Vector Machines (SVM) and Message Passing Neural Networks (MPN). Based on the detection confidence, the system labels the event as high or medium risk and sends an alert. A location-based SMS feature is included to notify nearby police stations, helping speed up emergency response.
Although the system performs well—achieving over 85% accuracy—it faces some challenges. Privacy concerns arise due to continuous audio monitoring, and background noise or variations in scream characteristics can affect accuracy. The SMS alert mechanism is also still under development. Despite these limitations, the study shows strong potential for using audio-based machine learning to improve public safety through faster detection of distress signals and quicker law enforcement intervention.

3. METHODOLOGY

3.1. Data Collection

The dataset for this system consists of two primary audio categories: human screams and non-scream environmental sounds. Positive samples include various types of distress sounds, while negative samples contain everyday noises such as talking, traffic, music, and background ambience. These audio files are gathered from publicly available datasets and manually recorded samples to ensure diversity. The data is organized into structured folders, and unsuitable files—such as extremely short clips or stereo recordings—are filtered out to maintain consistency for training the machine learning models.

3.2. Preprocessing and Feature Extraction

Before training, each audio file undergoes preprocessing to enhance quality and ensure uniformity. The signals are converted to mono, normalized, and trimmed to remove unwanted silence. The system then extracts Mel-Frequency Cepstral Coefficients (MFCCs), which capture essential frequency and time-domain characteristics unique to human screams. Each audio sample is transformed into a fixed-size MFCC matrix, allowing the models to learn recognizable acoustic patterns despite variations in pitch, loudness, or duration. This feature extraction process forms the foundation of accurate scream recognition.

3.3. Model Training

Two machine learning models are trained using the extracted MFCC features. The first is a Convolutional Neural Network (CNN), which processes the MFCCs as images and learns deep patterns associated with screams through multiple convolution and pooling layers. The second model, a Support Vector Machine (SVM), acts as a secondary classifier using statistical audio features. Training both models enhances the system's reliability by combining the strengths of deep learning, which excels at complex pattern recognition, with traditional machine learning, which offers interpretable and fast classification. Both models achieve high accuracy after repeated training and validation cycles.

3.4. Dual-Model Detection

During real-time operation, the system captures incoming audio and processes it using the same MFCC extraction method applied during training. The CNN predicts the probability that the audio contains a scream, while the SVM provides an additional verification. The outputs from both models are analyzed together to improve decision accuracy and reduce false alarms. This dual-model approach ensures that detections are more robust, especially in noisy or unpredictable environments where single-model systems may struggle.

3.5. Risk Level Classification

After obtaining predictions from the CNN and SVM, the system evaluates the confidence scores and classifies the detection into three risk levels: high, medium, or low. High-risk alerts are generated when both models strongly indicate the presence of a scream, while medium-risk alerts represent moderate confidence levels. Low-risk results typically correspond to normal environmental sounds or model uncertainty. This step enables the system to prioritize alerts and avoid unnecessary notifications, ensuring that only genuine emergencies receive immediate attention.

3.6. Location Retrieval

For significant detections classified as high or medium risk, the system automatically retrieves the approximate geographic location of the user. It utilizes IP-based geolocation to determine latitude, longitude, and the nearest city or address. This location data allows emergency responders or designated contacts to identify the user's position quickly, even if the person cannot communicate their location verbally. The inclusion of location data greatly enhances the usefulness of the alerting system.

3.7. Automated SMS Alerting

Once a high- or medium-risk scream is confirmed, the system triggers an SMS alert using the Twilio API. The message includes details such as the risk level, model confidence, time of detection, and the user's location or Google Maps link. These alerts are sent instantly to emergency contacts or local authorities, enabling rapid intervention. The automated SMS mechanism ensures that potential emergencies are communicated even when the victim is unable to reach out for help manually.

3.8. User Interface Deployment

To make the system accessible and easy to operate, a user-friendly interface is developed using the Kivy framework. The application allows users to record live audio, upload files for testing, and view detection results in real time. It also manages the background processes for location retrieval and SMS alerts. This interface integrates all system components into a functional platform suitable for real-world deployment in homes, public spaces, or personal safety applications.

4. WORKFLOW

4.1 Audio Input Acquisition:- The workflow begins with the system capturing audio either through a built-in microphone for real-time monitoring or by allowing users to upload an existing audio file. The system continuously listens to the surrounding environment and collects sound data without requiring manual intervention. This ensures that any sudden scream or distress signal is immediately captured. The recording process is optimized to handle different acoustic environments, making the system suitable for both indoor and outdoor scenarios. By providing seamless audio acquisition, the system forms the foundation for effective scream detection.

4.2 Preprocessing of Audio Signal:-Once the audio is collected, the system performs preprocessing to ensure that the input is clean and consistent. The signal is converted to a mono channel and normalized to maintain uniform amplitude levels. Noise reduction techniques are applied to minimize background disturbances that may interfere with detection accuracy. The audio is trimmed or padded to match the required duration, ensuring that all inputs meet the standardized length for feature extraction. This preprocessing step helps the model focus on the true characteristics of the audio rather than unwanted noise or variations.

4.3 Feature Extraction Using MFCCs:- After preprocessing, the audio undergoes feature extraction using Mel-Frequency Cepstral Coefficients (MFCCs). MFCCs effectively capture the frequency and time-domain patterns that are unique to human screams. Each audio sample is converted into a fixed-size MFCC representation (40 × 174 matrix), which serves as the input for the machine learning models. This step transforms raw audio into numerical features that highlight key acoustic signatures such as pitch, intensity, and spectral shape. MFCC extraction ensures that the system can analyze screams with high precision even in noisy or unpredictable environments.

4.4 Dual-Model Classification (CNN + SVM):- The extracted features are then passed through two separate machine learning models: a Convolutional Neural Network (CNN) and a Support Vector Machine (SVM). The CNN analyzes the MFCCs as if they were images and produces a probability score indicating how likely the audio is to contain a scream. The SVM serves as a secondary classifier that verifies the detection using statistical feature patterns. By combining the outputs of both models, the system significantly reduces false positives and improves overall detection reliability. This dual-model approach leverages both deep learning and traditional machine learning for robust classification.

4.5 Risk Assessment and Categorization:-Once the models produce their predictions, the system evaluates the confidence scores and categorizes the event into predefined risk levels. High-risk alerts indicate strong evidence of a scream with high model confidence, while medium-risk alerts represent uncertain or borderline cases that may require user confirmation. Low-risk outputs are generally associated with normal ambient sounds or unclear audio signals. These risk categories help prioritize alerts and reduce unnecessary notifications. The assessment mechanism ensures accurate decision-making, especially in sensitive scenarios where false alarms could cause panic or unnecessary intervention.

4.6 Location Retrieval (For High/Medium Risk)
If the detected event is classified as high or medium risk, the system automatically retrieves the user's location to support emergency response. Using IP-based geolocation, the system identifies the approximate latitude, longitude, and address of the device. A Google Maps link is also generated for quick access by emergency responders. This step enables authorities or contacts to locate the user even when they cannot verbally communicate their location. The integration of real-time geolocation improves the practicality and usefulness of the system in real-world emergencies.

4.7 Automated SMS Alerting:- Based on the risk level and user confirmation (if required), the system triggers an SMS alert using Twilio's communication API. The alert message includes the type of risk detected, the confidence level of the classification, the time of detection, and the user's location. These details enable emergency contacts or law enforcement officials to assess the situation immediately and take prompt action. The automated nature of the SMS system ensures that individuals receive help even if they are unable to call for assistance themselves. This step adds a crucial safety layer to the overall system design.

4.8 User Interface Deployment:- To ensure usability, the entire workflow is integrated into a user-friendly graphical interface developed using the Kivy framework. The interface allows users to record live audio, upload files for testing, and view detection results along with risk levels. It also manages background processes such as location retrieval and SMS alerting. The UI simplifies system interaction and enables practical deployment in various environments, including homes, public spaces, schools, and personal safety applications.

4.9 Continuous Monitoring and Feedback:- After completing a detection cycle, the system immediately returns to listening mode, allowing continuous monitoring without manual resets. The system logs detection events, risk levels, and alert outcomes, enabling future analysis and performance improvements. Continuous monitoring ensures that every potential distress signal is captured promptly, supporting proactive safety and quick emergency response.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

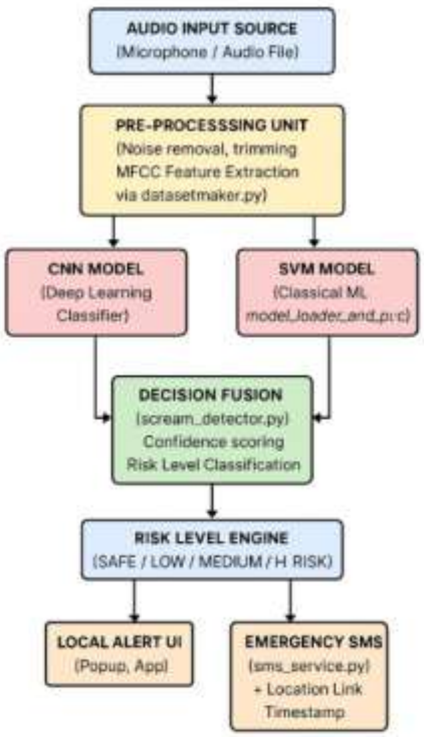

Fig 4.1 Work Flow Diagram

5. RESULT AND DISCUSSION

The scream detection system performed well during testing, with both the CNN and SVM models achieving more than 92% accuracy in identifying scream and non-scream sounds. The system was able to detect screams reliably in real-time, and the dual-model approach helped reduce false alarms by confirming detections through both models. The risk classification feature correctly separated detections into high, medium, and low levels, with high-risk cases triggering automatic SMS alerts and location sharing. Although the system worked effectively, noisy environments and very soft screams sometimes lowered confidence levels, and SMS delivery depended on network conditions. Overall, the results show that the system is accurate, practical, and useful for improving public safety, with potential for further enhancement in real-world applications.

6. CONCLUSION

This project successfully developed a real-time scream detection system using machine learning techniques, specifically CNN and SVM models, supported by MFCC-based feature extraction. The system demonstrated strong accuracy in distinguishing human screams from background noise and proved effective in real-time scenarios. The integration of dual-model verification improved reliability by reducing false alarms, while the risk classification module enabled clear decision-making for emergency responses. The system also incorporated automated SMS alerts and location tracking, providing a practical safety solution that can support faster intervention during critical situations.

Although the system performed reliably, certain limitations such as environmental noise, variations in scream characteristics, and network dependency for SMS alerts still exist. Addressing these challenges can further enhance system performance. Future improvements may include expanding the dataset, using more advanced deep learning models, strengthening noise-handling techniques, and improving real-time processing capabilities. Overall, the proposed system shows strong potential for real-world deployment in public places, homes, and personal safety applications, contributing significantly to proactive crime prevention and emergency response.

REFERENCE

[1] "Identification of Illicit Activities & Scream Detection using Computer Vision & Deep Learning" R.Mathur, T. Chintala, D. Rajeswari
Department of Data Science and Business Systems, SRM IST, Tamil Nadu, India. (2022)

[2] "Detecting Voice Features for Criminal Case"Aie Su Su Ky , K Zin Lin
University of Information Technology, Yangon, Myanmar (2021)

[3] "An Expert System for Identification of Domestic Emergency based on Normal and Abnormal Sound",Swati Shilaskar; Shripad Bhatlawande; Aditya Vaishale; Prapti Duddalwar; Aditya Ingale ,
Dept. of E&TC Engineering, Vishwakarma Institute of Technology, Pune (2023)

[4] "Human Scream Detection and Analysis for Crime Reduction", Srimathi A; Jothi S; Sindhiya M
Dept of CSE, KCG College of Technology, Chennai (2024)

[5] "Scream and Gunshot Detection and Localization for Audio-Surveillance Systems", Valenzise G.; Gerosa L.; Tagliasacchi M.; Antonacci F.; Sarti A.
Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy. (2007)

[6] "Two-Stage Supervised Learning-Based Method to Detect Screams and Cries in Urban Environments"M. Malhotra; A. Singh; A. Kumar
Indraprastha Institute of Information Technology (IIIT-Delhi), India. (2018)

[7] . "DcaseNet: An Integrated Pretrained Deep Neural Network for Acoustic Scene and Event Detection" J. Jung; H. Heo; J. Kim
Korea University, Seoul, South Korea. (2020)

[8] "Compact Recurrent Neural Networks for Acoustic Event Detection on Low-Energy Devices" Cerutti G.; Lightstone R.; Finn S.
Philips Research, Eindhoven, Netherlands. (2020)

[9] "Sound Event Detection and Time-Frequency Segmentation from Weakly Labelled Data" Qiuqiang Kong; Yin Cao; Turab Iqbal; Mark D. Plumbley
Centre for Vision, Speech and Signal Processing, University of Surrey, UK. (2018)

[10] "DEW: A Wavelet Approach for Rare Sound Event Detection"Bugeau A.; Legrand Y.; Le Moan S.
Centre Borelli, ENS Paris-Saclay, France. (2024)

[11] "Novel Sound Event and Sound Activity Detection Framework Based on Intrinsic Mode Functions and Deep Learning"Dattatraya G.; U. Rajendra; A. Singh
Dept. of Electronics Engineering, VNIT Nagpur, India. (2024)

[12] "Sound Event Detection in Traffic Scenes Based on Graph Convolutional Networks"Wang Z.; Zhang L.; Peng Y.
State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, China. (2024)

[13] "Detection of Emergency Situations Using Deep Learning-Based Sound Recognition Methods" Bose R.; Das A.; Mukherjee S.
Dept. of Computer Science, Techno India University, Kolkata, India. (2022)

[14] "ScreamSense: Enhancing Emergency Response with Deep Learning"Shashikala S.; Srinivas H.
Dept. of Computer Science, PES University, Karnataka, India. (2025)

[15] "Audio-Based Crime Prevention System Using Real-Time Scream Detection"R. Patel; M. S. Chauhan; J. D. Patel
Dept. of CSE, GTU, Gujarat, India. (2025)