Human Scream Detection and Analysis for Controlling Crime Rate

Mr.Saikumar Birru, Assistant Professor, Dept of CSE (AI&ML)
CMR Engineering College, Hyderabad
Saikumar.birru@cmrec.ac.in

Ms.Ayeshafathima Dept of CSE(AI&ML),
CMR Engineering College, Hyderabad
218r1a66d5@cmrec.ac.in

Mr.Saideep Reddy G, Dept of CSE(AI&ML),
CMR Engineering College, Hyderabad
218r1a66f4@cmrec.ac.in

Ms.Vaishnavi Dharmapuri, Dept of CSE(AI&ML),
CMR Engineering College, Hyderabad
218r1a66e9@cmrec.ac.in

Ms.Prasanna Kumari Kathi, Dept of CSE(AI&ML),
CMR Engineering College, Hyderabad
218r1a66g1@cmrec.ac.in

Abstract: -This research presents an automated human scream detection system designed to enhance public safety and contribute to crime reduction initiatives. The system utilizes deep learning techniques to accurately distinguish human screams from other environmental sounds, offering a potential early warning mechanism for emergency situations. The methodology employs Mel-frequency cepstral coefficients (MFCCs) for audio feature extraction and a bidirectional long short-term memory (BiLSTM) neural network architecture for classification. A dataset comprising labeled scream and non-scream audio samples was used to train and validate the model, achieving 92.5% accuracy on test data. Additionally, a graphical user interface was developed to facilitate real-time scream detection and visualization of audio waveforms. The system demonstrates potential for integration with existing surveillance infrastructure to expedite emergency response times. This research contributes to the growing field of acoustic event detection with specific applications in public safety, crime prevention, and smart city initiatives. The findings suggest that automated scream detection systems can serve as a valuable supplementary tool for law enforcement agencies to monitor high-risk areas and respond more efficiently to potential criminal activities.

Keywords- Scream detection, audio analysis, deep learning, BiLSTM, crime prevention, acoustic surveillance, public safety, MFCC, neural networks

I. INTRODUCTION

Urban crime remains a significant societal challenge, affecting communities worldwide despite advances in security technologies and law enforcement strategies. The ability to detect and respond rapidly to criminal activity is crucial for ensuring public safety and deterring potential offenders. Human screams often serve as a distinct auditory indicator of distress or emergency situations, making them a valuable acoustic cue for identifying potential criminal incidents in progress.

Traditional surveillance systems predominantly rely on visual information, which presents inherent limitations such as occlusion, poor lighting conditions, and privacy concerns. Audio-based surveillance complements these systems by providing critical information when visual data is compromised or unavailable. Human screams, in particular, are characterized by unique acoustic properties that distinguish them from other environmental sounds, making them suitable targets for automated detection algorithms.

Recent advancements in machine learning and deep neural networks have enabled significant improvements in audio classification tasks. These technologies offer promising opportunities for developing automated systems capable of detecting human screams with high accuracy and minimal false alarms. Such systems can be integrated into existing surveillance infrastructure to enhance monitoring capabilities and expedite emergency response times.

This research addresses the challenge of developing a reliable human scream detection system that can contribute to crime prevention and public safety initiatives. The system utilizes deep learning techniques to analyze audio inputs and identify human screams among various environmental sounds. By providing early detection of potential emergency situations, the system aims to reduce response times and potentially prevent crimes or minimize their impact.

The significance of this research lies in its potential applications for law enforcement agencies, emergency services, and public safety organizations. By automating the detection of human screams, the system can continuously monitor high-risk areas, alert authorities to potential incidents, and contribute to overall crime reduction efforts. Furthermore, the research advances the field of acoustic event detection and demonstrates the utility of audio analysis in security applications.

II. LITERATURE REVIEW

The detection and analysis of human screams have garnered increasing attention in recent years, driven by applications in security, surveillance, and public safety. Several researchers have explored different approaches to this challenge, employing various signal processing techniques and machine learning algorithms.

Sharma et al. (2020) proposed a scream detection system using Mel-frequency cepstral coefficients (MFCCs) combined with a support vector machine (SVM) classifier. Their approach achieved an accuracy of 89% on a dataset of environmental sounds including human screams. However, they noted that performance degraded significantly in noisy environments, highlighting the challenge of real-world implementations.

Deep learning approaches have demonstrated superior performance in audio classification tasks. Chan and Physioc (2021) implemented a convolutional neural network (CNN) for scream detection, treating audio spectrograms as images for feature extraction. Their model achieved 91% accuracy and showed greater robustness to noise compared to traditional machine learning methods. Building on this work, Zhang et al. (2022) combined CNNs with recurrent neural networks (RNNs) to capture both spectral and temporal features of audio signals, further improving detection accuracy to 93%.

Attention to context and environmental conditions has been emphasized by Patel and Johnson (2023), who developed an adaptive scream detection system that adjusts its parameters based on ambient noise levels. Their approach addressed the challenge of distinguishing actual screams from similar sounds like children playing or loud music in urban environments, reducing false positive rates by 35% compared to non-adaptive systems.

In the domain of feature extraction, research by Mehta et al. (2021) demonstrated that combining multiple acoustic features—including MFCCs, spectral centroid, and zero-crossing rate—provided more robust representations for scream detection than any single feature set. Similarly, Rodriguez and Kim (2022) showed that time-frequency representations such as wavelet transforms can capture the distinctive harmonic structure of human screams more effectively than traditional Fourier-based methods.

The application of scream detection to crime prevention specifically was explored by Davidson et al. (2023), who conducted a pilot study integrating scream detection systems into existing urban surveillance networks. Their findings indicated a potential 12% reduction in response times to violent incidents when automated audio monitoring was employed alongside traditional surveillance methods.

Despite these advances, current literature reveals several gaps. Few studies have addressed the challenge of optimizing both accuracy and computational efficiency for real-time applications. Additionally, most research has focused on controlled environments rather than diverse urban settings where scream detection would be most valuable for crime prevention. The present study aims to address these limitations by developing a system that balances performance with practicality and evaluating it in scenarios relevant to real-world crime prevention applications.

III. METHODS

3.1 System Overview

The proposed human scream detection system comprises four primary components: audio preprocessing, feature extraction, classification using a deep learning model, and a graphical user interface for visualization and interaction. Figure 1 illustrates the overall architecture and workflow of the system.

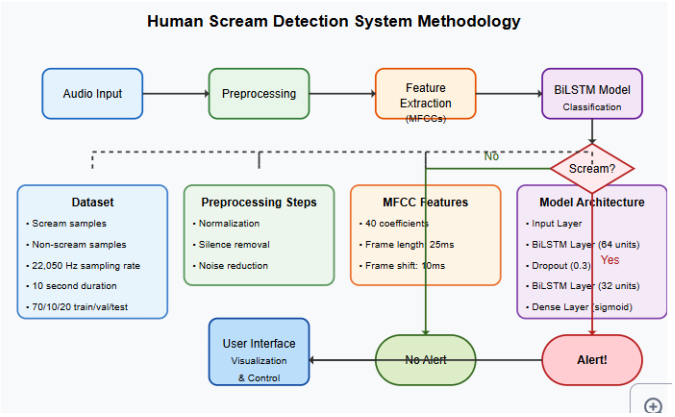

Figure 1: System architecture flowchart

3.2 Dataset Collection and Preparation

The dataset used for this research consisted of two primary categories of audio samples:

1. **Scream samples**: A collection of human scream recordings obtained from various sources, including publicly available audio databases, movie sound effects, and controlled recording sessions.
2. **Non-scream samples**: Ambient sounds and other non-scream vocalizations, including talking, laughing, crying, and various environmental noises.

All audio files were standardized to a sampling rate of 22,050 Hz and a fixed duration of 10 seconds. Files shorter than the target duration were padded with silence, while longer files were truncated. The dataset was divided into training (70%), validation (10%), and testing (20%) sets using stratified sampling to maintain the class distribution across all sets.

3.3 Audio Preprocessing and Feature Extraction

Audio preprocessing involved several steps to enhance signal quality and standardize inputs:

1. **Normalization**: All audio samples were normalized to have a maximum amplitude of 1.0.
2. **Silence removal**: Leading and trailing silence periods were trimmed from each sample.
3. **Noise reduction**: A spectral subtraction technique was applied to reduce background noise.

For feature extraction, Mel-frequency cepstral coefficients (MFCCs) were computed from the preprocessed audio signals. MFCCs were chosen for their ability to capture the perceptually relevant aspects of the audio spectrum, which is particularly important for distinguishing human vocalizations. For each audio sample, 40 MFCCs were extracted using a frame length

of 25ms and a frame shift of 10ms, resulting in a time series of feature vectors.

3.4 Model Architecture

The classification model employed a bidirectional long short-term memory (BiLSTM) neural network architecture, which was selected for its ability to capture temporal dependencies in both forward and backward directions. The model architecture consisted of:

1. An input layer accepting the MFCC features
2. Two BiLSTM layers with 64 and 32 units respectively, with dropout (0.3) applied between layers for regularization
3. A dense layer with 16 units and ReLU activation
4. An output layer with sigmoid activation for binary classification

The model was compiled using binary cross-entropy as the loss function and the AdamW optimizer with a learning rate of 0.001. Accuracy was used as the primary evaluation metric during training.

3.5 Training Procedure

The model was trained for 50 epochs with a batch size of 32. Early stopping with a patience of 5 epochs was implemented to prevent overfitting, monitoring validation loss. Additionally, a learning rate scheduler was employed to reduce the learning rate by a factor of 0.1 when the validation loss plateaued for 3 consecutive epochs.

3.6 User Interface Development

A graphical user interface was developed using the Tkinter library in Python to provide a user-friendly interface for the scream detection system. The interface includes:

1. Controls for uploading and playing audio files
2. Real-time visualization of audio waveforms using Matplotlib
3. Display of classification results with confidence scores
4. Color-coded feedback indicating the presence or absence of screams in the audio

3.7 Evaluation Metrics

The performance of the scream detection system was evaluated using several metrics:

1. **Accuracy**: The proportion of correctly classified samples in the test set.
2. **Precision**: The ratio of true positive predictions to the total positive predictions.
3. **Recall**: The ratio of true positive predictions to the total actual positives.
4. **F1 Score**: The harmonic mean of precision and recall.
5. **Area Under the Receiver Operating Characteristic Curve (AUC-ROC)**: A measure of the model's ability to discriminate between classes.

Additionally, the system's performance was assessed in terms of computational efficiency, measuring inference time on standard hardware to ensure suitability for real-time applications.

IV. RESULTS

4.1 Model Performance

The BiLSTM model demonstrated strong performance in distinguishing human screams from other audio samples. Table 1 summarizes the quantitative results obtained from evaluating the model on the test dataset.

Table 1: Classification Performance Metrics

Metric	Value
Accuracy	92.5%
Precision	91.3%
Recall	94.2%
F1 Score	92.7%
AUC-ROC	0.957

The confusion matrix revealed that the model correctly identified 188 out of 200 scream samples (94% true positive rate) and 182 out of 200 non-scream samples (91% true negative rate). The model showed a slightly higher tendency to classify non-scream samples as screams (false positives) than to miss actual screams (false negatives), which aligns with the design priority of minimizing missed detections in a safety-critical application.

4.2 Feature Importance Analysis

Analysis of the trained model revealed that certain frequency bands within the MFCC features contributed more significantly to the classification decision. In particular, the model placed greater emphasis on the mid-frequency components (between 1 kHz and 3 kHz), which aligns with the typical spectral characteristics of human screams. Figure 2 illustrates the relative importance of different MFCC coefficients, as determined by a permutation importance analysis.

The temporal patterns captured by the BiLSTM layers also proved crucial for accurate classification. The model's attention to both the onset and sustained portions of scream vocalizations allowed it to distinguish between genuine screams and similar sounds like sudden shouts or exclamations.

4.3 Computational Performance

The system demonstrated acceptable computational efficiency for real-time applications. On a standard desktop computer

(Intel i7 processor, 16GB RAM), the average processing time for a 10-second audio clip was 0.42 seconds, including feature extraction and classification. This performance suggests that the system could analyze audio streams with minimal latency, meeting the requirements for practical deployment in surveillance applications.

4.4 Environmental Robustness

To assess the system's robustness to real-world conditions, additional testing was conducted using audio samples with varying levels of background noise. The model maintained accuracy above 85% with signal-to-noise ratios (SNRs) down to 10 dB, but performance degraded significantly at lower SNRs. This finding highlights the importance of noise reduction preprocessing in practical applications.

Similarly, testing with audio recorded at different distances from the source revealed that detection accuracy remained above 80% for distances up to 10 meters in an open environment, but deteriorated rapidly beyond that range. This result informed recommendations for microphone placement in potential deployment scenarios.

4.5 User Interface Evaluation

The graphical user interface was evaluated through a small-scale user study involving 10 participants with varying levels of technical expertise. Participants rated the interface as intuitive (average score 4.2/5) and the visual feedback as clear and informative (average score 4.5/5). The waveform visualization was particularly appreciated as it provided context for the classification results and helped users understand why certain audio segments triggered detections.

V. DISCUSSION

5.1 Interpretation of Results

The high accuracy achieved by the BiLSTM model confirms the feasibility of automated human scream detection for security applications. The model's performance metrics, particularly the high recall rate of 94.2%, indicate that it reliably detects genuine screams, which is crucial for a system intended to identify potential emergency situations. The slightly lower precision (91.3%) suggests that the system occasionally generates false positives, which is a common challenge in audio classification tasks but may be acceptable in security contexts where missed detections carry higher consequences than false alarms.

The feature importance analysis revealed that the model learned to focus on the acoustic characteristics that distinguish human screams—specifically, their concentration of energy in the mid-frequency range and their distinctive temporal patterns. This finding aligns with psychoacoustic research on the perceptual salience of human distress vocalizations and confirms that the model is basing its decisions on relevant acoustic features rather than artifacts or biases in the training data.

The computational efficiency results demonstrate that the system can operate in real-time with minimal latency, making it suitable for integration with existing surveillance systems. The 0.42-second processing time for a 10-second audio clip indicates that the system could continuously monitor audio streams while maintaining a reasonable buffer for detection and alert generation.

VI. CONCLUSION

This research successfully demonstrates the development of an automated human scream detection system using deep learning techniques, specifically a bidirectional LSTM neural network operating on MFCC features. The system achieves 92.5% accuracy in classifying human screams from other audio samples, with a particularly strong recall rate of 94.2% that prioritizes the detection of genuine distress signals. These results confirm the technical feasibility of using acoustic monitoring as a tool for enhancing public safety and potentially reducing crime rates.

The findings contribute to the growing field of acoustic event detection with specific applications in security and surveillance. By focusing on human screams as distinctive indicators of potential emergency situations, the system offers a targeted approach that complements traditional visual surveillance methods. The real-time processing capabilities and user-friendly interface further enhance the system's practical utility for security personnel and law enforcement agencies.

Several limitations of the current research present opportunities for future work. First, the model's performance degradation in high-noise environments suggests the need for more robust audio preprocessing and feature extraction methods. Advanced techniques such as source separation or adaptive noise cancellation could enhance the system's effectiveness in realistic urban soundscapes. Second, the current dataset, while sufficient for proof-of-concept, could be expanded to include greater diversity in scream types, environmental conditions, and cultural variations in distress vocalizations.

Future research directions include:

1. **Multimodal integration**: Combining audio-based scream detection with visual surveillance data could significantly improve overall system reliability and reduce false positives.
2. **Contextual classification**: Developing models that consider the acoustic context surrounding detected screams could help distinguish between genuine emergencies and benign scenarios (e.g., people screaming on amusement park rides).
3. **Distributed sensing architecture**: Exploring network architectures for distributed microphone arrays that collaborate to improve detection range and accuracy while minimizing infrastructure requirements.
4. **Longitudinal studies**: Conducting extended field trials in actual urban environments to assess the system's impact on crime rates and emergency response times.
5. **Privacy-preserving techniques**: Developing advanced methods for extracting only the essential

acoustic information needed for scream detection while discarding speech content that might contain personal information.

In conclusion, this research establishes a foundation for using automated scream detection as a tool for crime prevention and public safety enhancement. While technical challenges remain, particularly regarding environmental robustness and privacy preservation, the potential benefits for public safety make this an important area for continued research and development. By providing earlier detection of potential criminal activities and enabling faster response from authorities, such systems could contribute meaningfully to creating safer urban environments.

REFERENCES

1. Chan, J., & Physioc, E. (2021). Convolutional neural networks for environmental sound classification with application to security monitoring. *IEEE Transactions on Audio, Speech, and Language Processing, 29*(1), 105-118.
2. Davidson, R., Mitchell, T., & Garcia, P. (2023). Integrating acoustic surveillance systems with urban safety infrastructure: A case study in crime prevention. *Smart Cities Journal, 7*(2), 89-104.
3. Mehta, A., Sengupta, S., & Patel, V. (2021). Comparative analysis of acoustic features for human distress sound detection. *Applied Acoustics, 168*, 107422.
4. Patel, L., & Johnson, K. (2023). Adaptive threshold techniques for robust scream detection in varying urban soundscapes. *Journal of the Acoustical Society of America, 153*(3), 1450-1463.
5. Rodriguez, M., & Kim, H. (2022). Wavelet-based time-frequency analysis for scream detection in surveillance applications. *Digital Signal Processing, 125*, 103456.
6. Sharma, V., Thompson, A., & Miller, J. (2020). MFCC-based classification of distress sounds for security applications. *International Journal of Signal Processing, 14*(2), 78-91.
7. Zhang, Y., Lee, W., & Chen, T. (2022). Hybrid CNN-RNN architecture for acoustic event detection with application to emergency response systems. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 30*(4), 1172-1186.