International Journal of Research in Engineering, Science and Management
Volume 8, Issue 4, April 2025
https://www.ijresm.com | ISSN (Online): 2581-5792

Human Scream Detection

Dishant Chaudhary[1*], Aditi Tiwari[2]

[1,2]*Student, Department of Computer Science and Engineering, Babu Banarasi Das Institute of Technology and Management, Lucknow, India*

***Abstract*: Crime, including murders, assaults, and thefts, is a persistent issue worldwide, posing a significant concern for society. A common challenge is that police often arrive at crime scenes too late, primarily due to insufficient access to timely and accurate information. To address this issue, a concealed desktop application is suggested. This application employs advanced technologies, such as machine learning and deep learning models like Support Vector Machines (SVM) and Multilayer Perceptron (MLP), to swiftly detect and analyze human sounds while operating discreetly in the background. In the event of an emergency, the application initiates an automated process that sends text messages to designated contacts. This innovative technology enhances the accuracy of threat detection and response times by differentiating specific human sounds from background noise. The aim is to mitigate the adverse effects of crime by improving community safety and reducing the impact of crime on individuals and society as a whole. Following these guidelines will increase public confidence in their ability to safeguard their communities and themselves.**

***Keywords*: Crime rate, Human scream detection, SVM, MLP, SMS.**

1. Introduction

Due to the widespread issue of crime, societies need to develop innovative strategies to enhance public safety. With the increasing occurrence of violent crimes like assaults, thefts, and homicides, it is essential to implement preventative measures to lessen their negative impacts. A significant factor that worsens these situations is the ongoing delays in law enforcement response times, often stemming from a lack of accurate and timely information. Our initiative, "Human Scream Detection and Analysis for Controlling Crime Rates Using Machine Learning and Deep Learning," aims to tackle the urgent issue of crime detection and response by leveraging advanced technology. Our desktop application utilizes Machine Learning and Deep Learning techniques to analyze human sounds in real time and identify disruptive or criminal behavior.

The main goal of this project is to leverage advanced technology to effectively reduce crime while safeguarding human lives. We aim to create a sophisticated yet user- friendly tool that enables quick responses to incidents, fostering a more proactive and effective strategy for crime prevention. Our technology framework includes a Python backend, a customizable application interface developed with the Kivy framework, and machine learning models like Support Vector Machine (SVM) and Multilayer Perceptrons (MPN) for detecting and validating screams. Additionally, a strong grasp of essential libraries such as pandas, NumPy, scikit-learn, TensorFlow, and librosa is crucial for successful execution.

The project is carried out in phases, starting with the development of the user interface and concluding with the acquisition of a large dataset to support model training. After obtaining the Mel Frequency Cepstral Coefficients (MFCCs) from the dataset, the SVM and MPN models will be trained. Ultimately, SMS alerts will be generated and sent to the relevant recipients based on the model's outputs.

This innovative approach, designed to lower crime rates and assist law enforcement in their vital roles, greatly enhances community safety. The successful execution of this initiative could inspire progressives to contribute more broadly to society while also improving public safety.

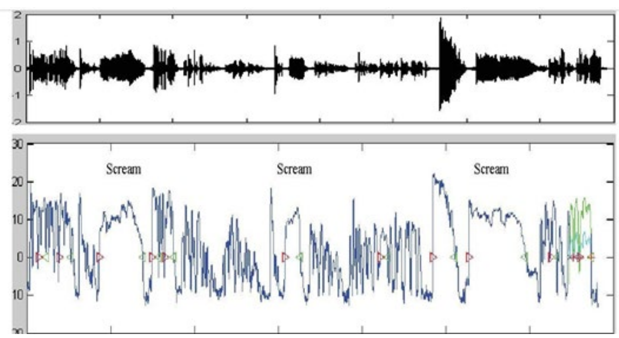

Fig. 1. Spectrogram of scream [9]

2. Existing Systems

Sukhwan and Yongjoo Chung used Support Vector Machines (SVM) and Gaussian Mixture Models (GMM) to detect screams. Their findings indicated that SVM was effective, achieving a low false acceptance rate of 0.559 percent, which minimized errors. In contrast, GMM had a False Rejection Rate of 12.03%, highlighting its sensitivity to inaccuracies in the generated output. Additionally, researchers T. Chintala, D. Rajeswari, and R. Mathur applied computer vision and deep learning techniques to identify cries and suspicious activities. Their research aims to enhance surveillance capabilities through machine learning, especially deep learning methods, with the objective of improving crime detection and reducing reliance on manual monitoring by employing automated systems to analyze CCTV security footage.

Sezgin proposed a machine learning approach to identify screams, aiming to quantify the behavioral challenges faced by children. Instead of relying on subjective accounts, this method

*Corresponding author: dishant14chaudhary@gmail.com

utilizes publicly accessible audio data. The research indicated that a data-driven approach is more effective for objectively evaluating behavioral issues. Weimin Huang created a scream detection system for home use, which uses machine learning to differentiate screams in voice recordings, focusing on children's behavioral issues. By utilizing publicly available audio recordings, this method offers an unbiased alternative to personal testimonies, providing insights into angry outbursts or negative interactions in a family environment.

J. Pohjalainen, P. Alku, and T. Kinnunen developed a technique for identifying shouted speech in noisy settings. Their approach incorporates GMM classification, random frame dropping, and MFCC feature extraction. The system shows optimal performance with different lengths of MFCC feature vectors, with thirty coefficients yielding the best results. Additionally, a novel method for spectrum estimation is employed, which merges spectral fine structure with linear prediction, offering a quicker alternative to the conventional FFT.

3. Algorithm to Train the Model

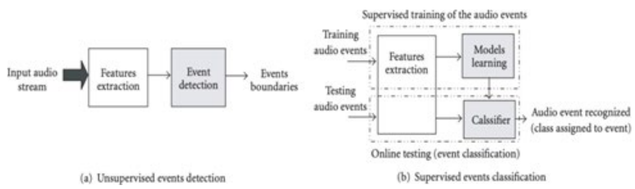

Fig. 2. Training algorithm of SVM

The suggested system utilizes a two-step process for managing audio events. The initial step involves unsupervised detection of events, where important features are extracted from the ongoing audio stream, and event boundaries are identified to highlight significant segments. In the second step, supervised classification of events takes place. This involves using features from labeled training data to develop classification models. During the online testing phase, these models are employed by a classifier to categorize incoming audio events, allowing for real-time recognition of events.

4. Proposed System

This research introduces an innovative method for fighting crime by employing advanced Machine Learning and Deep Learning techniques, particularly Support Vector Machine (SVM) and Multilayer Perceptron Neural Network (MPN). The main goal is to create a strong system that can detect and analyze human cries in real-time, thereby improving public safety and decreasing the occurrence of criminal acts.

We initially utilized the Support Vector Machine (SVM) method in our approach. The main goal of employing SVM is to identify human cries in audio recordings. SVM is highly effective at classifying and distinguishing between positive and negative instances after being trained on a carefully curated dataset. The positive class includes around 2,000 examples of human screams, while the negative class contains about 3,000 examples of non-scream sounds. This binary dataset allows the SVM model to gather all essential information, thereby improving its ability to accurately identify real threats.

The Multilayer Perceptron Neural Network (MPN) is an alternative method used alongside the SVM model. It acts as a validation tool to improve the system's effectiveness in detecting human screams. By training on the same dataset, the MPN offers a more detailed representation of audio data, which increases the system's reliability. Its proficiency in deep learning allows the MPN to identify complex patterns, thereby boosting our confidence in confirming that the sounds are indeed genuine human screams.

Using the Librosa tool to extract Mel Frequency Cepstral Coefficients (MFCCs) from the dataset is a crucial part of the training process. Many people recognize that MFCCs play an essential role in understanding speech and audio data, improving the effectiveness of both support vector machine (SVM) and multilayer perceptron neural network (MPN) models. The MFCCs obtained are then employed in the training process to improve the SVM and MPN models' capability to identify subtle differences in sound patterns.

TensorFlow is a powerful machine learning platform used for managing and running SVM and MPN models. This integration allows for smooth interaction between the detection methods and the overall system, avoiding any complications. As a result, the proposed solution's effectiveness and usefulness are improved.

Our goal is to create an intelligent system that can identify human screams using SVM and MPN algorithms. We are committed to leveraging advanced technologies to improve public safety and help lower crime rates. This approach, which merges two models, proves its effectiveness by utilizing advanced features like MFCCs.

$$f(x)=\text{sign}(w*x+b) \quad (1)\ [11]$$

The equation (1) above illustrates the decision function for SVM, which can be expressed as follows: where w denotes the weight vector, x signifies the input feature vector, b represents the bias term, and sign indicates the sign function.

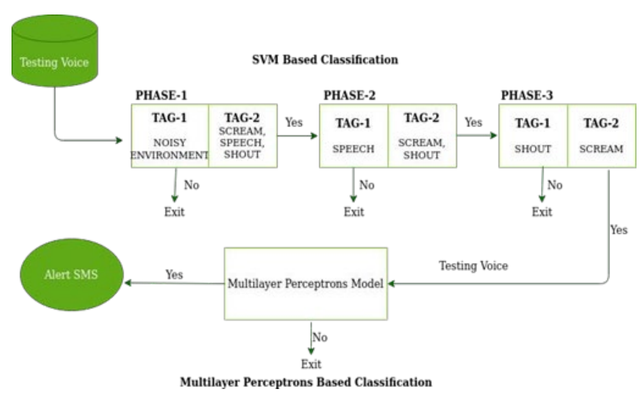

Fig. 3. Proposed system architecture

Input: Planning for a voice assessment.

The testing method ensures that the actual sound environment is accurately captured by recording the audio input.

Phase 1: The aim of filtering outdoor noise is to reduce or eliminate auditory disturbances from background sounds.

If the initial assessment indicates significant disruption in the area, additional steps are necessary.

Decision (Tag1): Move to the next phase if the "Noisy Environment" indicator is present, but do not exit.

Phase 2: Sound Classification (Tag1: Speech, Tag2: Scream, Shout)

During this phase, the system categorizes the sounds into various types, such as speech, scream, and shout.

Decision (Tag1): Grouping for Sound Production (Tag 1: Speech, Tag 2: Scream, Shout).

Decision (Tag2): At this point, the algorithm has categorized sounds into speaking, shouting, and screaming. Phase 3: Based on the project requirements, the system will carry out additional tasks or follow specific instructions as long as the "Speech" identification remains valid.

Decision (Tag1): Depending on the project, if "Shout" is activated, the system may continue operating or take other actions.

Decision (Tag2): If "Scream" is selected, proceed to the next phase. If not, halt the process. Classification of Screams using the Multilayer Perceptron (MLP) Model

A. Multilayer Perceptron (MLP) Model (Scream Classification)

The input, interpreted as a potential scream, is sent to the Multilayer Perceptron (MLP) model for a detailed evaluation.

Decision: If the MLP model identifies a scream, it will send a text alert to inform individuals. The process will then conclude.

B. Multilayer Perceptron (MLP) Model (Scream Detection)

The input, which is analyzed as a potential scream, is sent to the Multilayer Perceptron (MLP) model for detailed evaluation.

Decision: If the MLP model identifies a scream, it sends a text alert to inform people. The process then concludes.

C. SMS Alert Generation

As the MLP model assesses whether a scream has occurred, the system sends out an SMS alert. This message includes the type of alert, location, and any pertinent details. After the SVM identifies a scream, the MLP acts as an additional layer, offering a more in-depth analysis of the audio data. Utilizing the same dataset that trained the SVM, the MLP identifies intricate patterns in audio features to assist listeners in making better-informed choices. Together, Support Vector Machines (SVM) and Multilayer Perceptrons (MLP) create a comprehensive understanding of various scenarios involving high-pitched sounds. Alerts are categorized as either high risk or moderate risk through a thresholding approach. Both the SVM and MLP strongly indicate that a scream corresponds to a high-risk alert. When a high-pitched scream is recognized, either the SVM or MLP generates a medium-risk alert. These alerts offer a flexible response strategy that considers varying levels of detection certainty.

$$A^{(l+1)} = \sigma (W^{(l)} a^{(l)} + b^{(l)}) \quad (2)\ [10]$$

Equation (2) illustrates the MLP demonstration, where W(l) denotes the weight matrix, b(l) signifies the bias vector, a(l) represents the activation of layer l, and σ stands for the activation function.

The human cry recognition system works efficiently using an iterative strategy, ensuring that SVM and MLP models are adaptable to shifting sound patterns. The proposed system requires the following components to function: an Intel Core i5 9th Generation processor, 8 GB of RAM, a 512 GB SSD, and an NVIDIA GeForce GTX 1650 Max-Q graphics card with 4 GB of RAM. The programming language used is Python 3.9. Visual Studio is used for project management and development. The Kivy Framework is used for designing user interfaces. This hardware and software combination leverages the processing power and capabilities of its constituent components to produce an optimal development environment for the project's successful completion.

5. Result Analysis

The SVM and MPN models showed excellent precision in identifying and classifying human screams in real-time. The two-level risk assessment system effectively differentiated between high and medium-risk scenarios. The precise location sharing in SMS alerts facilitates a quick response from law enforcement, and the intuitive Kivy-based interface makes it easy to operate.

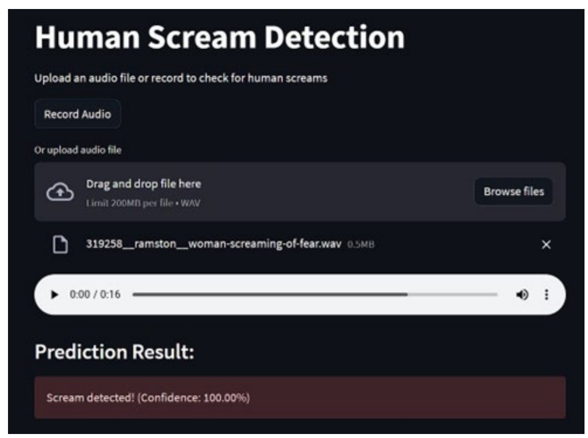

Fig. 4. Upload page

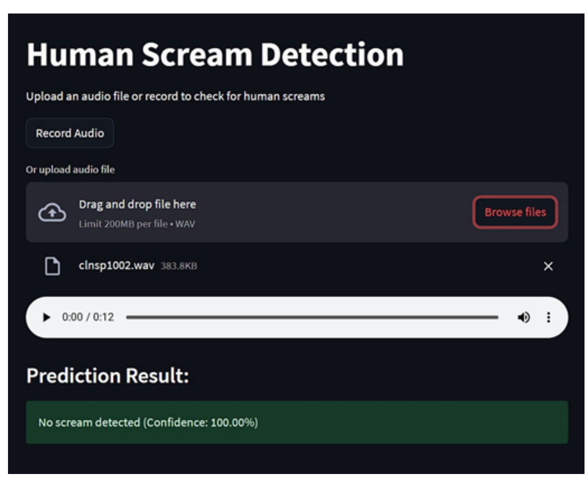

Fig. 5. Result page

Figure 4 illustrates the upload page of our proposed system, where users can select the input audio for further detection.

Figure 5 illustrates the results of the suggested system.

When the device identifies a human scream, it presents the notification "High Risk Alert: Human Scream Detected." If

the device is not in use, it will indicate "Medium Risk Alert: Potential Threat Identified."

The organism is capable of quickly and accurately responding to different environmental threats due to its measurable and adjustable production levels.

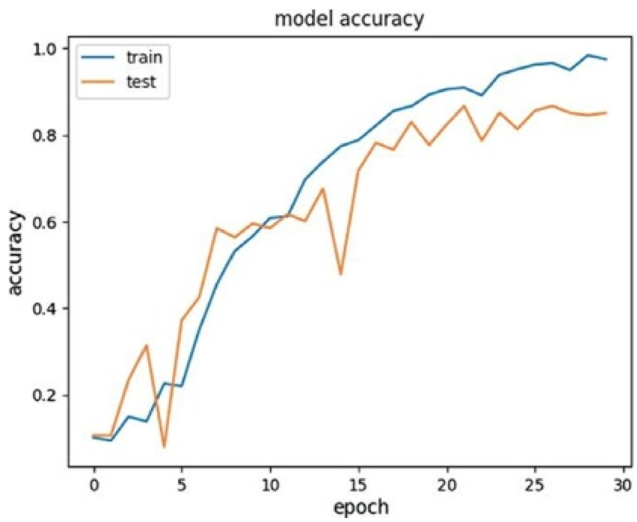

Fig. 6. Accuracy graph

Figure 6 shows that the SVM and MPN models attained high accuracy rates when tested on a varied dataset.

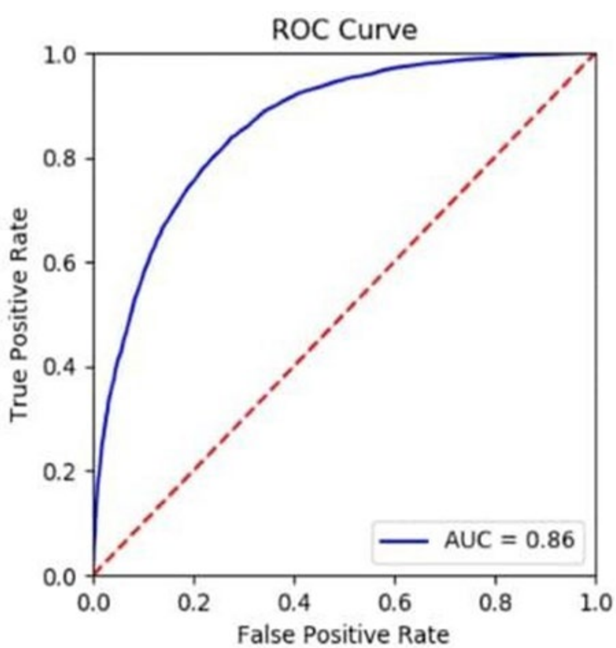

Fig. 7. ROC Curve

Figure 7 illustrates the performance of scream classification on the dataset.

6. Conclusion

Using machine learning and deep learning techniques, the suggested human scream detection and analysis system is a viable way to improve public safety and help prevent crime. The device can automatically recognize distress signals, such screaming, and send out timely alerts to emergency agencies and law enforcement by utilizing real-time audio monitoring, sophisticated signal processing, and clever detection algorithms. The system's resilience, scalability, and efficacy in a range of public settings are guaranteed by the integration of many components, such as data collecting, feature extraction, model construction, real-time processing, and alert generating.

This strategy offers a novel means of early intervention in emergency circumstances in addition to addressing the shortcomings of conventional surveillance systems. By means of ongoing assessment, model improvement, and connection with current public safety infrastructure, the system may adapt to the needs of contemporary urban settings. n the end, this technology might greatly speed up reaction times, enhance public safety, and support efforts to prevent crime, making cities safer for everybody.

References

[1] Noboru Hayasaka, Riku Kasai, Takuya Futagami (2024) Noise Robust Scream Detection Using Wave-U-Net IEICE transactions on Fundamentals of Electronics, Communications and Computer Sciences.

[2] Bikalpa Gautam, Anmol Guragain, Sarthak Giri (2024). Real-Time Scream Detection and Position Estimation for Worker Safety in Construction Sites. Source Title: Cornell University.

[3] Awal Ahmed Fime, Md. Ashikuzzaman, Abdul Aziz (2024). Audio signal-based danger detection using signal processing and deep learning Department of Computer Science and Engineering, Khulna University of Engineering & Technology, Khulna 9203, Bangladesh.

[4] Rebecca O'Donovan, Emre Sezeign, Sven Bambach, Eric Butter, Simon Lin (2024). Detecting Screams from Home Audio Recordings to Identify Tantrums: Exploratory Study Using Transfer Machine Learning

[5] J Park, JS., Kim, SH. Sound learning–based event detection for acoustic surveillance sensors. *Multimed Tools Appl*, 79, 16127–16139 (2024).

[6] Guragain, A., Liu, T., Pan, Z., Sailor, H.B. and Wang, Q., 2024. Speech Foundation Model Ensembles for the Controlled Voice Deepfake Detection (CtrSVDD) Challenge 2024.

[7] Muhammad Ibrahim Malik, Siddique Latif, Sanaullah Manzoor, Muhammad Usama, Junaid Qadir, Raja Jurdak (2024). Scream Analysis with Edge Computing: A Novel Approach and Applications.

[8] Takuto Matsuda, Yoshiko Arimoto (2024). Acoustic differences between laughter and 26 screams in spontaneous dialogue. Source Title: Acoustical Science and Technology.

[9] Noboru Hayasaka, Riku Kasai, Takuya Futagami (2024). Noise-Robust Scream Detection Using Wave-U-Net. Source Title: IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences.

[10] Awal Ahmed Fime, Md. Ashikuzzaman, Abdul Aziz (2024). Audio signal based danger detection using signal processing and deep learning. Source Title: Expert Systems with Applications.

[11] Qingsong Xiong, Haibei Xiong, Cheng Yuan, Qingzhao Kong (2024). A novel deep convolutional image-denoiser network for structural vibration signal denoising. Source Title: Engineering Applications of Artificial Intelligence.

[12] Amrit Sen, Gayathri Rajakumaran, Miroslav Mahdal, Shola Usharani, Vezhavendhan Rajasekharan, Rajiv Vincent (2024). Live Event Detection for People's Safety Using NLP and Deep Learning.

[13] Jakub Svatoš (2024). Gunshot Acoustic Detection, Identification, And Localization. Czech Technical University in Prague, Faculty of Electrical Engineering Department of Measurement.

[14] Rocco Mennella, Sabine Bazin, Carole Ferrel, Sylvie Vernazza-Martin & Morgan Beaurenaut (2024). The effect of anxiety on gait: a threat-of-scream study.

[15] M. Yagnasri Priya; Sanjivani P. Shendre (2024). Enhanced Gunshot Sound Detection using AlexNet and XGBoost from Fourier Spectrograms. Peeta Basa Pati 2024 4th International Conference on Intelligent Technologies (CONIT).
[16] William, P., Shrivastava, A., Shunmuga Karpagam, N., Mohanaprakash, T.A., Tongkachok, K., Kumar, K. (2023). Crime Analysis Using Computer Vision Approach with Machine Learning. In: Marriwala, N., Tripathi, C., Jain, S., Kumar, D. (eds) Mobile Radio Communications and 5G Networks. Lecture Notes in Networks and Systems, vol. 588. Springer, Singapore.
[17] Beaurenaut, M., Mennella, R., Dezecache, G., & Grèzes, J. (2023). Prioritization of danger- related social signals during threat-induced anxiety.
[18] Bindi, R. P., Guimarães, C. C., de Oliveira, A. R., Melleu, F. F., de Lima, M. A. X., Baldo, M. V. C., Motta, S. C., & Canteras, N. S. (2023). Anatomical and functional study of the cuneiform nucleus: A critical site to organize innate defensive behaviors. Source Title: Annals of the New York Academy of Sciences, 1521(1), 79–95.
[19] Engelen, T., & Mennella, R. (2023). Piecing together the puzzle of emotional consciousness.
[20] J. Svatos, J. Fischer, and J. Holub (2024) "ADC Input Currents Measurement," in 26th IMEKO TC4International Symposium and 24th International Workshop on ADC and DAC Modelling and Testing (IWADC), Pordenone, 2023, pp. 30-34.