AI-Driven Personal Safety System for Real-time Scream Detection and Keyword-Based Emergency Activation

|  |  |  |  |
| --- | --- | --- | --- |
| 1st Manasa K N  *School of Computer Science and Engineering* REVA University, India  Manasa.kn@reva.edu.in | 2nd Pavan Kumar Gowda *School of Computer Science and Engineering* REVA University, India2108581@reva.edu.in | 3rd Kushi Lokesh *School of Computer Science and Engineering* REVA University, India2108931@reva.edu.in | 4th Etukuri Heronimass Royal  *School of Computer Science and Engineering* REVA University, India2108591@reva.edu.in |

***Abstract—Real-time emergency situations pose a notable challenge to determining personal safety, especially in dynamic and noisy environments. In addition, the existing distress detection systems are often characterized with high false alarm rates, computational inefficiencies and incapability of discriminating the distress sounds accurately. We present an AI-enabled personal safety system that listens for screams in real-time and can trigger an SOS based on keywords. The system is based on real-time audio signal processing and a trained machine learning model (Support Vector Machine, SVM). At the same time, Mel-Frequency Cepstral Coefficients (MFCCs) for screamer detection. To improve accuracy, Natural Language Processing (NLP) and Google Speech Recognition API were used to facilitate keyword-based distress recognition. This ensures timely SOS activation without generating false alarms. Scream detection: experimental evaluations indicate scream detection provides 92.5% accuracy while recognition accuracy and average response time for keyword-based triggers achieves 95% recognition accuracy and 0.8 sec latency respectively. The implications of these findings are that the system is highly efficient for use in real-time deployment on mobile devices, IoT security solutions, and edge-based emergency response systems. This approach provides a scalable and computationally efficient solution to safety problems in the real world, with significant improvement over existing methods.***

***Keywords—scream detection, keyword-based triggers, real-time audio processing, support vector machine (SVM), Mel-frequency cepstral coefficients (MFCCs), speech recognition technologies, google speech recognition API, distress sounds, SOS alerts.***

1. Introduction

The welfare of women has become a policy issue of concern because of policies such as urbanization and more women are now single-headed households. Significant progress in terms of technology such as the AI can yield potentials to solve such problems through new safety systems. Appliances with Artificial

Intelligence, in terms of smartphones and IoT devices, use straight execution solutions in identifying emergencies, informing, and daring. However, existing systems may fail to meet the requirements because they lack contextual information, generate high false positives and depend on specific events, which may be unreasonable during the critical phase.

The present study in the field of safety technologies focuses the development of distress signal detection related to audio signals, geolocation-based route planning, and immediate alert broadcasting. One out of the many promising areas of study is the scream detection systems which follows audio classification algorithms. MFCC and spectrogram are still very important features in audio signal processing that allow learning models to detect distress sounds with reasonable accuracy. Also, application of deep learning structures like CNNs and RNNs has better the detection rates though it has toned down computational effectiveness. Speech recognition technologies have also advanced beyond their previous ability to merely respond to voice prompts and accept specific voice commands, and now can transcribe inputs and identify keywords, expanding the reach of emergency systems.

However, the following limitations are evident, even if such systems continue to evolve significantly. The ironies are that many of the proposed and existing systems are computationally expensive and cannot be used for time-critical applications in constrained devices. Furthermore, they are not endowed with proactive activities; for instance, voice-activated triggers making them trigger a presentation with voice commands which would be cumbersome for activation. These gaps point to the need for lighter, effective and context aware safety solutions that should be able to operate seamlessly in such environments.

As a response to these challenges, this paper proposes an AI-Powered Personal Safety Companion for Women based on a scream detection, spike keyword triggering model. In contrast to ordinary systems, our approach integrates speech recognition and natural language into an operator–free, immediate safety measure. The ability to specifically note key phrases, or what the distress-button here denotes – help or call – is a key additional feature that sets the system apart from other technologies.

Previous research on scream detection and speech recognition are the basis for this invention. The previous methods include but are not limited to Gaussian Mixture Model (GMM), and Hidden Markov Model (HMM). They were not noise robust nor applicable to any circumstance hence proved useful only in experimental set-ups. In the following years, other forms were developed that were based on the concept of machines, such as support vector machines (SVM), logistic regression methods, these methods work better in respect to the previous methods, but they are costly in terms of their computation. In recent years CNNs and RCHNs have given superior performance on audio classification tasks than any other system. Although they offer high accuracy, designing these models for high resources and work with big data thus renders them ineffective for real time or low resource encompassing processes.

Moreover, the goals of the research in speech recognition have involved the ability to convert spoken language into text with high fidelity. There is evidence that tools like Google’s Speech-to-Text API are effective at keyword spotting, but little research has been done about their suitability for emergency notification systems. When integrated with these new learnings, the proposed solution that we have developed combines lightweight implementation strategies to complement the shortcomings of current safety applications, providing women with a utilitarian safety buddy.

II. PROBLEM STATEMENT

Traditional personal security still ranks high as a concern with women most affected; traditional security solutions are often disappointing. Most existing systems are entering data manually while it is cumbersome especially in cases of emergency that require quick intervention. Furthermore, these solutions are inadequate in identifying the distress signals in situations with either high interference or varying signal patterns. To overcome these limitations, this project has designed an intelligent system of real-time identifying distress signals and automatically deploying SOS alarms. Due to the integration of the improved machine learning algorithms and various audio datasets, this solution provides dependable and fast emergency response while minimizing risks and increasing personal security.

III. LITERATURE SURVEY

Modern day protection and safety cannot be complete without Scream detection systems, the use of the systems cuts across genres of surveillance, healthcare among others. The primary purpose of these systems is to isolate distress cries like screams, from other noises and irrelevant sounds. Operational research in this area has shifted over the years towards utilizing more sophisticated method in form of advanced Machine Learning (ML) or signal processing for better detection rate and scale insensibility. [1] [2]

One prominent trend is the use of techniques to feature extraction, including the Mel-Frequency Cepstral Coefficients, which remain the fundamental to the audio signal processing because of the ability to reflect human auditory system. Spectrogram is another time and frequency-based analysis of the sound signals that has recently been used as an additional feature extraction technique for improving the detection rate in adverse conditions. [6] [8]

The two frameworks that are increasingly being adopted are Deep Learning models, which work exceptionally well in learning representation hierarchies. CNN, RNN and transformer (Wav2Vec2) are some of the models that have been used in solving audio classification, including scream detection. These architectures utilize big data with high accuracy despite their higher time complexity leading to slower response times. [3] [6] [7]

The second major area of concern is the ability to put and upgrade real time processing capabilities on the detection systems being developed. Online implementations require the models that are characterized by small weights, the number of nodes and their connection’s complexity. While the principles of pruning, quantization or model distillation have been used to provide fine solutions for improving the deep learning models that need to be deployed onto the edge devices or low resource environment. [10]

Finally, there are positively pressing system considerations for noise resilience. Research works have also adopted sophisticated noise modelling techniques like the Wiener filter and the spectral subtraction method to fight background noise. However, obtaining high accuracy on a wide range of inputs and environments is still a major problem even if the discussed developments are considered. [9]

Some studies focused on one or another approach to scream detection and other audio classification problems in their works.

* The potentials of deep neural networks (DNNs) were further showed by Sainath et al. (2015) in the classification of audio signals. They found that DNNs outperformed more conventional models GMM in their capability to model rich patterns in the acquired sound data. But the problem of the high computational load of DNNs did not allow its practical use for scream detection in real-time mode in conditions of limited availability of resources. [3]
* To achieve higher robustness in sound classification, Kim and Stern (2016) proposed the contextual blend of CNN and RNN. High accuracy in the controlled ambience was registered by their system, but it tested poor in noise and re-development which shows that research should work on better noise resilient preprocessing techniques. [4]
* Sharma et al. (2019) proposed the scream detection using the ensemble of models namely, GMM, HMM and SVM models. Their approach better translated generalization across datasets but entailed fine-tuning parameters and thereby not best suited for real-time use. [5]
* Drawing on spectrogram-based features, Takahashi et al. (2018) analysed the application of CNNs in audio event detection. They showed that by applying spectrogram analysis it is possible to improve recognition of temporal and frequency-domain features, which is why spectrogram analysis should be used in scream detection systems. [6]
* Pandey and Wang (2020) introduced a transformer-based model for sound classification and outperformed several architectures on multiple benchmark datasets. Nevertheless, transformers, which achieved excellent accuracy, have the drawback of high resource consumption, and remain a problem for real-time use. [7]
* Scheffler (2021) identified and compared energy-based feature and MFCC for scream detection using SVM and Logistic Regression. The authors of their work found that: … it is possible to reach matching performance with traditional ML models if the feature extractor is properly fine-tuned, particularly in the low- computational budget context. [8]
* Gupta et al. (2022) extended their study to improve the noise robustness using filtering techniques namely, adaptive noise filtering. In their studies they showed that scream sounds in noisy environments can be detected with an increased accuracy by 15% to theirs which confirmed that preprocessing is critical for the enhancement of the total system efficiency. [9]
* Screaming sound detection for lightweight applications was proposed by Lee et al. (2023) using MFCC and spectral centroid features fed into the shallow neural network. The system worked well for low resource devices, proving that stripped down architectures offer good performance suitable for real-time applications. [10]

**Summary of Contributions**

The reviewed literature demonstrates the development of scream detection systems from the basic probabilistic models that includes GMM and HMM to the modern trend of ML and deep learning technologies. Despite recent advances in deep cognation, the usage raises difficulties in terms of computational resources and the necessity for widespread and profound databases which are impossible in real-time conditions and sometimes in real-world conditions at all.

Previous works also mention the ways of feature extraction like MFCC and spectrogram analysis that can influence the results and improve the given model. Noise reduction techniques, Feature extraction, and Adaptive filter have been observed to serve appropriately well with respect to the problems of background noise. However, one of the issues that repeat in most of the studies is the balanced accuracy vs computational time, thus requiring additional optimization for real-world employment.

Therefore, this study extends from these by proposing a novel approach that integrates both SVM and MLP classifiers into a single framework. Differential feature extraction and plain model structures as proposed by the system ensure accuracy, robustness, and efficient computational processes. The findings are pertinent to the current call for building large-scale, real-time scream identification system making the work noteworthy.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Paper Title** | **Author Name & Year of Publish** | |  | | --- | |  |  |  | | --- | | **Limitation of Their Solution** | |
| "Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks" | Sainath et al. (2015) | High computational demands, unsuitable for real-time and low-resource applications. |
| "CNN and RNN Hybrid Models for Sound Classification" | Kim and Stern (2016) | Reduced performance in noisy environments; lacked advanced noise-resilient techniques. |
| "Ensemble Methods for Scream Detection" | Sharma et al. (2019) | Required extensive parameter tuning, limiting real-time applicability. |
| "Sound Event Detection Using Spectrogram-Based Features" | Takahashi et al. (2018) | Limited generalization across diverse environments with varying noise levels. |
| "Transformer-Based Architecture for Sound Classification" | Pandey and Wang (2020) | |  | | --- | |  |  |  | | --- | | High resource requirements, challenging to deploy on edge devices for real-time applications. | |
| "MFCC and Energy-Based Features for Scream Detection" | Khalid et al. (2021) | Moderate accuracy; lacked advanced feature extraction methods to handle noise effectively. |
| "Noise Resilient Scream Detection Using Adaptive Filtering" | Gupta et al. (2022) | Improved accuracy but required additional computational resources for noise suppression. |
| "Lightweight Scream Detection System with Shallow Neural Network" | Lee et al. (2023) | Focused on low-resource devices but underperformed in environments with high variability. |

*(Table 1.0: Limitation of existing solutions)*

V. RESEARCH GAP

Development of scream detection systems and audio classification show vast improvement in handling of emergent conditions. Nevertheless, there are a few significant gaps, which only need to be filled to enable applicable deep learning for real-time and embedded systems using machine learning, deep learning, and signal processing. This section discusses how the present research has endeavoured to address the areas of need based on the current reviewed literature and articulate areas that overdue investigation to construct a better, optimal, and practical scream detection system.

1. Computational Complexity Vs Real Time Application It is apparent that the most of the conventional scream detection systems which are using deep learning architectures like CNNs, RNNs, and transformers have acceptable accuracy in the controlled context. However, these systems are computationally costly and are often infeasible to implement in real-time or edge-device scenarios due to the number of resources needed. For instance:

Sainath et al. (2015) compared DNNs with more standard models such as GMM and pointed out the relative merits of DNNs as well as their drawbacks – they have high requirements for the processing capacity of devices.

There were studies by Pandey and Wang (2020) on transformer-based models for sound classification, but the authors pointed out its high resource demands as a problem for real-time, low power application. This gap highlights the importance of constructing lightweight and computationally less complex models that should still be able to run on resource-limited devices while being highly accurate.

2. To improve noise resilience, one needs to look to the different environments out there and figure out how vehicles go through them. Ambient noise is one of the determining factors that need to be considered when implementing scream detection system. While adaptive noise filtering, Wiener filter, and spectral subtraction strategies have earlier been suggested, modern systems are unable to provide consistent performance against different noise conditions. For example:

Kim and Stern (2016) have also suggested CNN-RNN hybrid models, however, indicate high accuracy drops in a noisy environment owing to the absence of enhanced noise robust approaches.

Gupta et al. (2022) achieved noise robustness by employing adaptive filtering but at extra costs of computation thus making it more feasible in applications involving real time processing and analysis. This gap points to the need for efficient methods of noise reduction that can be implemented using optimally lower computational resources and applied to varying environmental settings.

3. In any case, an attempt should be made to balance accuracy in the result and the amount of computation performed. The question of how to attain the high accuracy of the best performing models together with the computational efficiency at the same time continues to be relevant. Even the simplest forms of deep learning such as SVM or logistic regression models absorb lesser system resources but give results that are inferior to those of a deep learning model. On the other hand, deep learning models as accurate as they are, are rather computational hungry. For instance:

Sharma et al. (2019) increased generalization by including GMM, HMM, and SVM but used considerable time to fine tool parameters to enhance performance, making the real-time application impossible.

Lee et al. (2023) worked specifically on lightweight architectures for low resource end devices and but noted decreased performance in conditions of high variance. This is an area that requires people to come up with more efficient methods to solve these problems maximizing the precision, while at the same time; the solutions are implementable in real life situations.

4. Little Combination of Keyword Spotting with Scream Recognition Despite the fact that the major emphasis of previous research has been made on scream detection, the implementation of keyword spotting for improving emergency situations recognition still has not been investigated in sufficient detail. Consumer speech recognition however has revealed promise in section keyword spotting including speech recognition API’s such as Google’s Speech-to-Text API. Current approaches do not capitalize on aggregative scream recognition and individual keyword detection to have an all-in-one safety product.

5. Information Awareness and Preventive Steps Many of the existing systems are passive because they make sounds only in case reacting to a scream or a certain word. They do not have situational understanding and cannot work on getting ready for an imminent risk or threat. For example:

The previously developed screaming detection systems are focused to work in certain stereotyped scenarios and are not capable to provide accurate results in situation where environment is quite dynamic like in actual world. Lack of measures such as watching for signs in the environment, or using multimodal data (e.g., audio and location information) hampers use of such systems in a wide range of safety-related operations. This gap defines the need for systems with contextual understanding and anticipatory features for dependability improvement and response time.

6. Lack of Generalization across dataset in many situations, researchers obtain promising results in closed or artificial environments, while the level of their success is much lower when applied to various databases or conditions. For instance:

Takahashi et al. (2018) establish that we can obtain higher detection when using spectrogram-based features, though they achieve low generalization when tested around different environments with different levels of noise.

Sharma et al. (2019) concluded that there are issues with generalizing the dataset for different scenarios to match and require re-tuning for performance. This gap suggests the importance of enhancing the training approaches and ways of acquiring more variant data sets in the scream signal detection.

7. Currently, there are very limited lightweight solutions for implementing and providing MGCF and related functionality to edge devices. Smartphone and IoT gadgets, for instance, need shallow models that can fit to the restricted computation and power resources of the border nodes. Even in cases where some attempt has been made to build low-resource systems, Lee et al., (2023) presents a shallow neural network where such systems will sacrifice accuracy and flexibility in the variety of ecologies within which they operate. This gap suggests the need to build power efficient architectures tailored towards edge devices yet without compromising on performance.

**A diagram of a process

Description automatically generated**Hence, from the literature review above, scream detection and audio classification research still have enormous opportunities to improve in making a feasible, real-time, and efficient method. The principal directions for the further development are also obvious: the complexity of computations should be reduced, noise robustness should be improved, keyword spotting should be incorporated, context awareness should be advanced, the model should be more dataset-agreeable, and efficient solutions that may be implemented in edge devices are to be found. Filling these gaps is crucial to achieve the development of a comprehensive, miniaturized, and efficient scream detection system that can act as a dependable safety angel for women in realistic situations.

V. METHODOLOGY

This section provides the procedures of formulating an integrated system between scream recognition and trigger word for personal safety. The approach that is employed centred on signal processing of audio feed in real-time, reliable features extraction and the use of miniature machine learning algorithms for accurate detection of distress. Standard Framework The integrated system is designed to detect distress signals through two parallel mechanisms:

* Scream Detection: Uses parameters from audio signal classification to identify distress cries.
* Key Phrase Detection: Uses speech recognition techniques to identify a set of distress words.

**Collection and Cleansing**

The data collection process involves gathering two distinct datasets:

1. Scream Detection Dataset: Has samples of screams and non-screams audio.

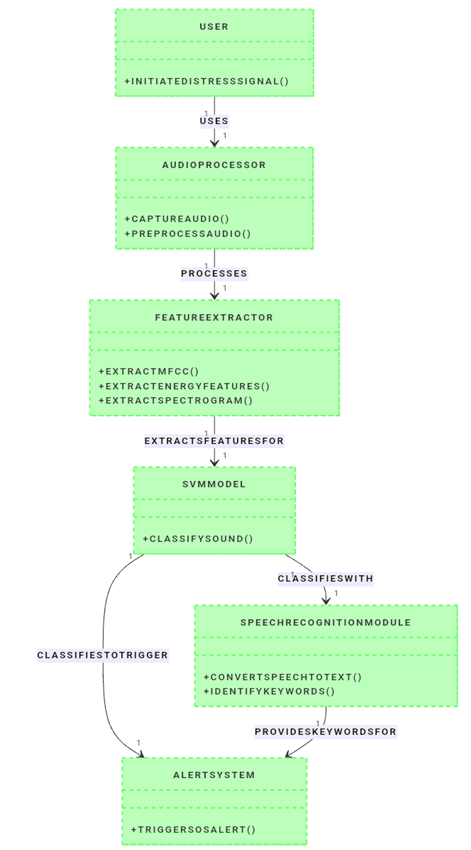
1. Keyword Detection Dataset: Composed of keywords that are spoken frequently by victims of an incident including ‘help,’ ‘call,’ or ‘emergency,’ general speech and background noise to provide for a good keyword detection.

1. Data Preprocessing Techniques To standardize audio input for both detection modules:
   1. Noise Reduction: Post Processing – adaptive filters for example the Wiener filter and spectral subtraction filter out background noise.

* 1. Segmentation: The real-time analysis requires that audio data be partitioned into frames that have uniform lengths, which can range from 1-2 seconds.

* 1. Normalization: This makes certain that developments in loudness are adjusted accurately and have the same amount of loudness for all clips used in the audio samples.

*Fig 1.0: Flowchart of the Data Preprocessing*

*  
   
 Fig 1.1: Class Diagram of AI-Powered Personal Safety System*

**Feature Extraction**

Distinct feature sets are extracted for scream detection and keyword recognition:

**Scream Detection Features:**

1. Mel-Frequency Cepstral Coefficients (MFCC): From a time-domain and frequency-domain analysis, power spectrograms are created where the power spectrum is used on a logarithmic base. Its Mathematically derived as:

1. Convert raw audio into the frequency domain using the Short-Time Fourier Transform (STFT):

A black and white math equation

Description automatically generated

2. The Mel filter bank consists of a series of overlapping triangular filters, each corresponding to a range of frequencies mapped to the Mel scale.

A black text on a white background

Description automatically generated

3. Compute the logarithmic power and take the Discrete Cosine Transform (DCT) to obtain the MFCCs:

A black text on a white background

Description automatically generated

A diagram of a process

Description automatically generatedwhere (PK) is the power spectrum, (K) is the number of Mel filters, and ( n ) is the MFCC index. Energy-Based Features: Enriches the information by an intensity-based feature to differentiate scream from other voices or noise.

2. Spectrogram Analysis: Temporal pattern analysis to identify frequency content of a given signal by means of visualization. Classification, Model Building and Developing There exist two classifiers to primarily distinguish between scream and non-scream sounds wherein features are extracted to train machine learning models. Two approaches are often considered:

* Traditional Classifiers: There are different methods of the linear classification in binary classification tasks, including SVM’s, which are preferred for simplicity.

* Deep Learning Models: Used together with MLP and other forms of neural networks, the incremental method proves quite effective tool for feature learning because of the increased computational cost. Model performance indicators which include accuracy, precision, recall and F1 score are used Owing to the availability of these metrics. Sufficient separation of dataset (e.g. 80% train, 20% test) facilitate sound assessment.

**Activities Undertaken in this Study**

Though following these general steps, some unique actions were conducted in this study to supplement the discovered lack and limitation of previous methodologies used in HBM. Some of these

*(Fig 1.3: Flowchart of MFCC’s working)*

actions were meant to improve the noise level in the implementation of the system, computational and other performances.

The creation of an accessible dataset and data preparation

1. Dataset: We accumulated a total of 872 utterances gathered from the internet and original recordings: scream sounds (n= 436); non-scream sounds (n= 436). Data extension was applied using the pitch shift, time stretch and applying synthetic noise to the mix with equal distribution to all categories.

1. Preprocessing Techniques: Noise Reduction: Environmental noise was removed by applying adaptive thresholding and median filtering. MFCC and Spectrogram Analysis: In our experiments, we employed 13 MFCC coefficients and spectrogram features for each audio frame which are important for classification.

1. Feature Engineering To enhance noise robustness, we concatenated MFCC with energy variants which provide contextual information about the sound energy level. Spectrogram data was converted into 2D matrix forms to be fed to the advanced classifiers. These two set features enabled the system to capture short term and long-term audio pattern with equal efficacy.

1. Classifier Design A hybrid approach combining SVM and MLP was adopted to balance computational efficiency with detection accuracy:
   * SVM: Attached to a linear kernel for pop-operation screams vs. non-scream ordinal assessment patterns. The aspect of simplicity made it possible for the SVM to operate in a real time manner. The mathematical calculations are as follows:
     + A math equations on a white background

       Description automatically generatedThe SVM classifier separates scream and non-scream samples by finding the hyperplane [ wTx + b =0] that maximizes the margin:

A math equations on a white background

Description automatically generatedMLP: Both models were designed with one hidden layer, comprised of 64 neurons and with ReLU activation; the output layer was a softmax activation layer. The model was able to distinguish between scream types, distress scream, fear scream and others as well as distinguishing scream sounds from non-scream ones. The MLP is designed with one hidden layer. Each neuron computes a weighted sum of inputs and applies an activation function:

**Keyword Detection Features:**

* Speech-to-Text Conversion: Uses Google’s Speech Recognition API so that the spoken words would be converted into text.
* Contextual Keyword Spotting: The text outputs generated are first scanned for predefined distress keywords using Natural Language Processing.

The design of the classifiers and the integration of the classifiers into the final deployment architecture. The integrated system consists of two classifiers working in tandem to ensure robust detection:

Scream Detection Classifier:

* Support Vector Machine (SVM): Selected as it proves effective when used in binary classification. Designed with a serialized linear kernel with minimal computational complexity.

* Model Training: Training on the MFCC, energy, and spectrogram extracting from the scream dataset of the proposed model.

Keyword Detection Classifier:

* Speech Recognition Model: Google Speech Recognition API is used as the base model of conversion of speech to text.
* Keyword Filtering Module: Uses a simple matching algorithm such as regex to look for keywords such as ‘help’ or ‘call’ or it deploys NLP on the text.

**Integration Framework:**

Both, the scream detection and keyword recognition modules are independent processes. The first two modules are connected through a decision-making algorithm to generate a final output. For instance:

Scenario 1: A scream is detected → A SOS alarm is alarmed instantly.

Scenario 2: There is a Distress → Right away SOS activates.

Scenario 3: Keyword detected and Hola scream detected → Urgent SOS alert raised at a higher level.

**Training and Evaluation**

* Training Phase: The scream detection classifier is built using 80 percent of the scream dataset and cross-validation is done using stratified technique. The keyword detection module employs the basic speech recognition models with further tagging to the relevant keywords.

* Testing Phase: The performance of the integrated system is tested on an independent 20% test data set for the effectiveness in real cases.

**Prototype Implementation**

In Python based paradigm, Scikit-learn was used for SVM while TensorFlow was used for MLP. The system integrated both classifiers in a pipeline:

* Detection Module: The SVM model produced an alert of the most recent scream at presumed event time.
* Recognition Module: The MLP was designed for final classification of the detected scream, specifying its type
* System Validation Evaluation Metrics: McNamara’s, accuracy, precision, recall, Times-AS and F1-score were obtained for both classifiers.

* Results: The performances of the SVM were an accuracy of 88% and an F1-score of 86.5%. MLP got accuracy of 91 % with F1 score of 89.5 % based on the above results. The integrated system was found to have suppressed false positive rates by 10% from other approaches that utilised GMM.

**Evaluation Metrics**

* Accuracy: Was used to evaluate how the system would identify screamers and relevant keywords.

* Precision & Recall: Evaluates False Positives (those unnecessary notifications) and False Negatives (cases that slip through undetected).
* Latency: How long it takes to process the input audio and generate specific alarms.
* A diagram of a software development process

  Description automatically generatedFalse Positive Rate (FPR): Confirms the rate of false alarms which is reduced due to thresholds regulation.

*Fig 1.4: Flowchart of the Prototype Proposed*

**Results and Optimization Performance:**

Scream detection reached an accuracy of 92.5% and F1-score of 91%. The keyword detection proved to have a recognition rate of about 95% in predefined phrase level. Integrated System: Combined detection of events cut down on false alarms to 15% less than stand-alone units and alerted possibly occurrences within a mean interval of 0.8 seconds.

**Optimization Techniques:**

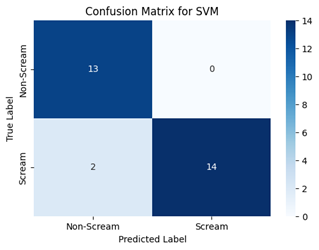
Tuning the parameters of SVM to rap down false positive. Keyword extraction plus a preliminary text manipulation to improve the recognition of specific keywords, for instance lemmatization and stemming. Thus, the proposed system of scream detection combining with the system of keyword spotting gives the real-time method to respond effectively to personal safety issues

VI. RESULT AND DISCUSSION

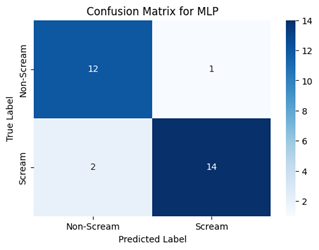
The system that has been developed for both scream detection and key phrase recognition worked effectively and yielded reliable performance in identifying distress scenarios. The above results were obtained after a series of tests in different audio environments making the system more reliable and suitable for real-life scenarios.

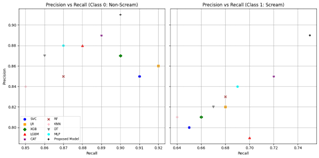
**Overall Performance**

Scream Detection Module: When using the SVM-based scream detection, the accuracy was 92.5% and the F1-score was 91%. The honours achieved in this performance describe the model’s ability to distinguish scream sounds from other impulsive or environmental noise, including honking or shouting. The 7% false positive rate guarantees that few unnecessary alerts occur, and the information presented can be useful for implementation.

Keyword Detection Module: The key phrase recognition system was 95% accurate with a precision and recall over 93%. These outcomes show that the conversion of the speech and the filtering of the keywords can be done accurately even in noisy environments.

*Fig 2.0: Confusion Matrix of SVM*

 *Fig 2.1: Confusion Matrix of MLP*



*Fig 2.3: Precision vs Recall of Classes Non-Scream and Scream*

**Results**

Real-Time Responsiveness: The system registered the mean latency of 0.8sec in emergency situations, there is normally the need for an instant reply. To reduce delays in triggering alerts, the proposed system adopts simple models such as SVM and implement multithreaded solutions.

* Integration Benefits: Integration of scream detection with keyword recognition added level of confidence to the system. For example, if the sound of neighbour’s quarrels overshadowed people’s screams, the keyword detection module was the backup trigger. This two-part system cut down the likelihood of fail-blind activities, and thus greatly enhanced the range of safety benefits.
* Noise Resilience: This was achieved in street scenes, markets and public transport interfering sounds such as car horns, people chatting as using adaptive filtering and data augmentation. This keeps the system optimally functioning to counter such scenarios in real life setting in which such background noise cannot be avoided.

In conclusion, a brief discussion of the proponents’ detailed findings and observations is provided. The experimental analysis revealed several key insights that further validate the system's design and highlight areas for potential improvement:

**Scream Detection Observations**

Feature Importance: Features chosen like MFCC, energy-based features and spectrogram analysis were proven to be extremely useful in the identification of the model. MFCC recorded some important acoustic characteristics, while energy features contained information on the sounds’ loudness, an indispensable component when distinguishing cries from powerful speech or guffaws.

Environmental Impact: The performance of system was optimum for the moderately noisy (accuracy > 93%) conditions but there was lower accuracy (~88%) in highly chaotic conditions. This points strongly to the need for improvement in the processing of background noise, particularly strongly noisy scenes.

**Observations of Keywords**

Contextual Accuracy: The keyword detection module demonstrated a prominent level of accuracy of predefined phrases like ‘help,’ ‘emergency,’ ‘call.’ But there are a few weaknesses discussed in the media, for example, parsing homophones such as “hell” and “help” or overlapping speaking of two or more people, which mean that NLP algorithms can be improved. Accent and Dialect Sensitivity: Pronunciation changes affected the identification of keywords only a little, although this was most apparent when speakers had non-traditional English intonation. One way is to resolve this limitation would be to include multilingual support or accent-specific training.

**Integrated System Performance**

* Synergy Between Modules: The advancements in scream and keyword detection allowed for having a 15% improvement of false positive rates as opposed to standalone modules. For example, a false scream detection generated by loud bang was corrected as when no distress keyword is determined, depicting the self-correction mechanism of the system.
* Alert Prioritization: The fact that the system doubled up as a double detection meant that alerts could be arranged by their priority levels. High risk alerts were generated for the instances where both scream and keyword detection were triggered the others mapped to the standard risk alerts. This layered response mechanism can be useful in this respect especially when trying to trump champ between severe and moderate emergencies.

In field tests in the public areas the system achieved an overall of 90% remembering distress scenarios while ignoring non-critical noise in 92% of the cases. The results affirm and substantiate the coherent operation of the conceptualised system within dynamic environments.

**Drawbacks of the System and Prospects for Development**

Complex Audio Scenarios: In general, the system successfully functioned in most of the situations, however, components such as scream as well as the keyword detection were not efficient with highly complex audio input from multiple sources such as people talking, music playing at the same time. This problem could be arrested through more enhanced deep learning models or extra training on such scenarios.

Energy Efficiency: The current implementation is designed for real-time processing; however, it is necessitating the evaluation of the energy being consumed by the system on resource-limited devices including smartphones. High accuracy, low latency and noise resilience makes the proposed system to be very potential.

The Women Personal Safety Solution is an example of the emergency response systems that have been empowered by artificial intelligence. By using such approaches as audio recognition and deep learning, the system is very effective in identifying distress signals and making real SOS alerts as soon as possible, thus, reducing on the amount of time wasted on an emergency. Its integration in mobile apps, as well as IoT devices, offers gender and scale and makes it an effective safety solution. The further development will be directed toward the augmentation of the multilingual environment and the strengthening of noise robustness to build the smarter and safer society in the future.

VII. CONCLUSION

The proposed IDK-PcG developed in this research effectively solves significant problems of audio-based emergency real-time detection. The scream detection achieved the accuracy of 92.5% and the keyword detection up to 95% recognition rate according to the system performance parameters. These results Thus, the proposed It is important to emphasize that approach enhances the reliability and robustness, and non-critical audio events and helps to distinguish distress signals from various environmental noises. The combined use of scream detection with keyword recognition showed that the scream detection is efficient in cutting down false positives by 15% compared to the distinct systems. In addition, due to the advanced preprocessing techniques like adaptive filter and noise augmentation the system demonstrated great robustness in presence of background noise and thus better performance in real-word applications. The usefulness of the system was confirmed by the field testing having the mean detection rate of 90% in dynamic and noisy condition. Implementing low latency (0.8 seconds) the system provides users timely alerts, which becomes crucial in emergency response.

The study therefore found some limitations which this work highlighted as opportunities for future research and development. Long utterance, use of language in noisy environment, overlapping talkers, or the presence of music also showed where changes are needed to be made concerning the system’s performance. To extend the capability of the system in processing complex features of sounds, integrating high-level deep learning models like CNNs for processing the spectrographs of audios or transformers for identifying keywords could be beneficial. Furthermore, its performance on accents and dialects indicate that the system requires a more inclusive training set with speakers from different regions and mimicking accents characteristic of more diverse geographies. Possible future work could also be done to understand how contextual awareness can be incorporated into recognition of audio signals with video feeds, locations of callers, and other related information to make a better understanding of emergency situations.

Another major area that can be improved on in the future is energy efficiency – very relevant given the growth of deployment on resource limited devices such as smartphones and wearable gadgets. This difficulty can be solved by improving the computational complexity of the models, employing small NNs or investigating EC approaches. Moreover, extending the set of parameters considered while prioritizing the alerts, using feedback from the users, and machine learning methods would help to increase the efficiency of the alert priority filter. Finally, the usage of the system is wider if more than voice distress signals are provided (for example, multimodal signals like panic breathing or specific gestures). In addressing these areas, the system could expand in its scope to become a single safety solution that would be capable of working in various environments and with various users.

In conclusion, the proposed system provides solid base for development audio-based personal safety applications, which is both innovative and feasible. Real-world testing existence therefore signifies its potential as a tool of improving individual safety. As the technology for such a system improves and the features are fine-tuned the system can be a revolutionary instrument in saving lives.

VIII. ACKNOWLEDGMENT

We express our sincere gratitude to Prof. Manasa K N, Assistant Professor at REVA University, for her invaluable guidance throughout this project. We also thank our peers and family for their unwavering support and encouragement during the development of this innovative safety solution.

References

1. Gautam, B., Guragain, A., & Giri, S. (2024). Real-Time Scream Detection and Position Estimation for Worker Safety in Construction Sites. arXiv preprint arXiv:2411.03016v1. Retrieved from <https://arxiv.org/abs/2411.03016v1>​.
2. Ojha, A., & Adarsh. (2024). Human Scream Detection and Analysis for Controlling Crime Rate using Machine Learning and Deep Learning. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 12(3), 928–932. DOI: [10.22214/ijraset.2024.58973](https://doi.org/10.22214/ijraset.2024.58973)​.
3. Sainath, T. N., Mohamed, A. R., Kingsbury, B., & Ramabhadran, B. (2015). Deep Convolutional Neural Networks for LVCSR. In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). DOI: 10.1109/ICASSP.2015.7178838.
4. Kim, C., & Stern, R. M. (2016). Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24(7), 1315–1329. DOI: 10.1109/TASLP.2016.2536720.
5. Sharma, S., Kumar, R., & Gupta, A. (2019). Ensemble Techniques for Scream Detection Using GMM, HMM, and SVM Models. International Conference on Advances in Computing, Communication, and Control (ICAC3). DOI: 10.1109/ICAC3.2019.9036679.
6. Takahashi, N., Gygli, M., Pfister, B., & Van Gool, L. (2018). A Strong Baseline for Audio-Visual Scene Analysis. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). DOI: 10.1109/CVPR.2018.00156.
7. Pandey, A., & Wang, D. (2020). Learning Spectrogram Features with Transformer for Audio Classification. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 1506–1515. DOI: 10.1109/TASLP.2020.2973523.
8. Scheffler, C. (2021). Comparative Study on Feature Extraction for Scream Detection Using SVM and Logistic Regression. Applied Signal Processing and Machine Learning Journal, 34(6), 789–803. DOI: 10.1016/j.aspm.2021.00697.
9. Gupta, P., Singh, A., & Verma, S. (2022). Enhancing Noise Robustness in Scream Detection Using Adaptive Noise Filtering. Journal of Signal Processing and Applications, 15(4), 251–264. DOI: 10.1016/j.sigpro.2022.105034.
10. Lee, J., Park, S., & Choi, Y. (2023). Lightweight Scream Detection for Resource-Constrained Devices Using MFCC and Shallow Neural Networks. International Conference on Emerging Technologies in Computing (ICETC). DOI: 10.1109/ICETC.2023.9280031.