International Journal of Engineering Research & Science (IJOER) ISSN: [2395-6992] [Vol-12, Issue-3, March 2026]

A Comparative Study of AI Techniques for Real-Time Personal Safety Application

Archie Patil[1*]; Pooja Narkar[2]; Pradnya Lohar[3]; Dr. Ashwini Save[4]

Department of Computer Engineering, VIVA Institute of Technology, Mumbai, India
*Corresponding Author

Abstract— *Personal safety has become an important concern in today's urban environments, especially for women and individuals who frequently travel alone. As safety concerns continue to evolve, academic research has explored a range of technology-based solutions aimed at enhancing personal security. A significant portion of this work investigates the role of artificial intelligence in understanding speech patterns, detecting emergency cues, and incorporating location-based information to assess risk. In this context, the present paper surveys research on AI-based personal safety systems, focusing on Speech Emotion Recognition, Trigger Word Detection, and unsafe-zone identification approaches. The surveyed literature is analyzed based on the types of models employed, speech features utilized, datasets referenced, and performance measures reported. A comparative analysis is included to outline observed performance differences, commonly adopted methodologies, and existing limitations across studies. Based on the survey, important research gaps are discussed, highlighting the need for integrated and context-aware safety systems that can operate reliably in real-time conditions. The findings of this review aim to assist future research efforts in designing more practical and automated AI-based personal safety frameworks.*

Keywords— AI Safety, Emergency Detection, Personal Security, Speech Emotion Recognition, Trigger Word Detection.

I. INTRODUCTION

Personal safety is increasingly becoming a major concern in urban areas, mainly due to rising incidents of harassment, theft, and unsafe travel experiences. Groups such as women, students, and individuals who travel alone are often more vulnerable, particularly when moving through unfamiliar or poorly monitored locations. To respond to these challenges, researchers have explored various technology-driven safety solutions, with a growing focus on mobile applications and AI-based monitoring systems [19].

Advances in artificial intelligence and deep learning have made it possible to design systems that can analyze human speech, emotional cues, and surrounding environmental conditions. **Speech Emotion Recognition (SER)** methods are commonly used to identify emotions such as fear, anger, or distress from voice signals, while **Trigger Word Detection (TWD)** techniques aim to recognize emergency-related keywords from continuous speech. SER systems have evolved from conventional classifiers to deep CNN–LSTM based architectures and semi-supervised models for improved robustness [16], [17]. Similarly, TWD approaches have progressed from template-based DTW methods [12] to more advanced alignment-free LF-MMI frameworks [20], multimodal audio-visual keyword spotting models such as KWS-Net [21], and efficient streaming transformer architectures for real-time wake-word detection [18]. Alongside speech-based approaches, location-aware safety systems utilize GPS information, crime data, and environmental factors to estimate risk levels in real time.

Despite extensive research in these areas, the existing literature shows considerable variation in model architectures, datasets, evaluation metrics, and practical applicability. Many proposed systems either depend on manual user interaction or focus on a single detection mechanism, which limits their effectiveness during real emergency situations. This highlights the need for a structured review that examines existing approaches collectively and identifies their limitations. Accordingly, this paper presents a literature survey of AI-powered personal safety systems and includes an analysis table comparing techniques, datasets, and reported outcomes. The aim is to summarize current research trends, highlight existing gaps, and provide insights that can support the development of more integrated, automated, and reliable personal safety solutions.

II. LITERATURE REVIEW

2.1 Speech Emotion Recognition Systems

Chaudhary et al. [1] proposed a speech emotion recognition (SER) system using machine learning techniques to classify human emotions from audio input. The system uses the RAVDESS dataset, which contains emotionally labeled speech

samples. Feature extraction is performed using the Librosa Python library, where speech signals are converted into spectrograms. Initially, a Random Forest classifier was used, but a Convolutional Neural Network (CNN) was later implemented, achieving significantly higher accuracy of 95%. For real-time implementation, the classified emotion is sent via Firebase to an ESP8266 Wi-Fi module, which lights up color-coded LEDs to indicate the detected emotion.

Sarker et al. [2] developed a text-independent SER system capable of classifying eight emotional states using solely audio input. The system extracts a joint feature set consisting of Mel-Frequency Cepstral Coefficients (MFCCs) and Log-Mel Spectrograms (LMS). These features are fed into a CNN for emotion classification. The model achieved an average accuracy of approximately 93% on the RAVDESS and TESS datasets, outperforming several existing models.

Waghmare et al. [3] presented an SER system using a Probabilistic Neural Network (PNN) to classify five basic emotions: happy, sad, angry, fear, and boredom. The system achieved 95.76% accuracy on EMO-DB and 84.64% on RAVDESS. The PNN model outperformed GMM, HMM, RNN, and CNN, demonstrating high reliability and simplicity.

Ouyang et al. [4] presented a deep learning system for speech emotion detection using MFCC and a hybrid CNN-LSTM architecture targeting seven emotions. Trained on SAVEE and RAVDESS datasets, the system achieved 61.07% accuracy, with best results for anger (75.31%) and neutral (71.70%).

Jena et al. [5] proposed a voice-based safety system that automatically detects distress using deep learning. The system monitors for the wake word "ON" and determines if it is spoken in a negative emotional tone. Using CNN for wake-word recognition and LSTM for emotion detection, it achieved 97.23% accuracy for the wake word and 88.94% for emotion detection on Google Speech Commands and RAVDESS datasets.

Franti et al. [6] developed a voice-based emotion recognition system for companion robots using CNNs. The system achieved a mean accuracy of 71.33% on Romanian speech data, demonstrating how visual CNN architectures can be adapted to audio spectrograms for real-time emotion classification.

2.2 Trigger Word Detection Systems

Zeng et al. [7] proposed a keyword spotting system combining DenseNet and BiLSTM to capture both local spectral features and long-range temporal dependencies. With only ~223K trainable parameters, the model achieved 96.6% accuracy on the Google Speech Commands dataset.

Bonet et al. [8] proposed a Speech Enhancement (SE) integrated Wake-Up-Word (WUW) detection system. Evaluated on a custom "OK Aura" dataset, the system improved accuracy from ~85% to ~93% under noisy conditions.

Kundu et al. [9] proposed HEiMDaL, a highly efficient CNN-based wake word detection system. Results showed a 73% reduction in false rejection rate compared to a DNN-HMM baseline at 12 false alarms per hour.

Kumar et al. [10] proposed a multi-task learning approach using hybrid CNN-LSTM for joint wake-word detection and text-dependent speaker verification, achieving ~30% relative improvement over DNN baselines.

Kumatani et al. [11] introduced a wake word detection system that directly processes raw audio signals using DNNs, achieving up to 12% AUC improvement over LFBE-based models.

Zehetner et al. [12] proposed a template-based wake-up word spotting system using Dynamic Time Warping (DTW) on MFCC features, achieving 99.7% precision with moderate recall (~59.6%).

2.3 Integrated Safety Systems

Priya et al. [13] developed the Safe Alert App, combining voice emotion detection (BERT, Wav2Vec 2.0) and visual threat recognition (YOLO + OpenCV), with LSTM and XGBoost for predicting weak network zones.

Usha Kumari et al. [14] introduced SOS, an audio-based safety app using CNN for sound classification and LSTM for emotional tracking, achieving over 85% accuracy on public and custom datasets.

Dole et al. [15] developed the WE SAFE App, featuring voice-based distress detection, GPS tracking, geofencing, and automatic alerts. Using XGBoost, Random Forest, and LightGBM on a 10,000-entry dataset, it achieved 93% accuracy with XGBoost and 92.5% with Random Forest.

III. ANALYSIS TABLE

TABLE 1
SUMMARY OF REVIEWED STUDIES ON AI-BASED PERSONAL SAFETY SYSTEMS

Sr. No.	Paper Title (Year)	Technology Used	Dataset	Key Result
Speech Emotion Detection				
1	Speech Emotion Recognition Using Neural Network (2020)	CNN, Random Forest, Librosa	RAVDESS	95% (CNN)
2	A Text-Independent Speech Emotion Recognition Based on CNN (2023)	CNN, MFCC + LMS	RAVDESS, TESS	93%
3	Application of Probabilistic Neural Network for SER (2024)	PNN	EMO-DB, RAVDESS	95.76% (EMO-DB), 84.64% (RAVDESS)
4	Speech Emotion Detection using MFCC and CNN-LSTM (2024)	MFCC, CNN + LSTM	RAVDESS, SAVEE	61.07%
5	Developing a Negative Speech Emotion Recognition Model for Safety Systems (2025)	CNN (wake word), LSTM (emotion)	Google Speech Commands, RAVDESS	Wake word: 97.23%, Emotion: 88.94%
6	Voice-Based Emotion Recognition for Companion Robots (2017)	CNN, Spectrogram	Romanian speech dataset	71.33%
Trigger Word Detection				
7	Effective Combination of DenseNet and BiLSTM for Keyword Spotting (2018)	DenseNet + BiLSTM	Google Speech Commands	96.60%
8	Speech Enhancement for Wake-Up-Word Detection (2021)	CNN + Denoising Autoencoder	Custom "OK Aura"	~85% → ~93% (noisy)
9	HEiMDaL: Detection and Localization of Wake-Words (2022)	CNN + Alignment Loss	500k augmented utterances	73% FRR reduction
10	Convolutional LSTM for Joint Wake-Word Detection and Speaker Verification (2018)	CNN-LSTM (multi-task)	WSJ, LibriSpeech	~30% improvement over DNN
11	Direct Modeling of Raw Audio with DNNs for Wake Word Detection (2017)	Raw Audio + DNN	Real-world Alexa data	~12% AUC improvement
12	Wake-Up Word Spotting for Mobile Systems (2014)	DTW on MFCC	Custom mobile dataset	Precision: 99.7%, Recall: ~59.6%
Integrated Safety Systems				
13	Safe Alert App (2025)	BERT, Wav2Vec 2.0, YOLO, LSTM, XGBoost	Custom/Real-time	High (not specified)
14	SOS: Bridging the Gap Between Threat and Safety for Women (2024)	CNN + LSTM	Public + Custom	>85%
15	WE SAFE App (2025)	CNN, LSTM, RF, XGBoost, LightGBM	10,000-entry dataset	XGBoost: 93%, RF: 92.5%

From the surveyed literature, it is evident that deep learning architectures such as CNN, LSTM, BiLSTM, and hybrid CNN–LSTM models consistently achieve stronger performance than traditional machine learning methods. Studies using MFCC and Mel-spectrogram features report higher accuracy and robustness, especially in real-world noisy environments.

IV. RESEARCH GAPS AND FUTURE DIRECTIONS

4.1 Key Research Gaps

Gap	Description
Gap 1	Most systems address safety components (SER, TWD, location) individually rather than as integrated solutions
Gap 2	Limited research on context-aware safety that combines emotional state, trigger words, and environmental risk simultaneously
Gap 3	Few systems have been validated in real-world emergency conditions (most rely on controlled datasets)
Gap 4	Lack of standardized benchmarks for evaluating personal safety systems across diverse acoustic environments
Gap 5	Privacy concerns regarding continuous audio monitoring remain largely unaddressed

4.2 Future Research Directions

1. **Integrated Multi-Modal Systems:** Combining SER, TWD, and location-based risk assessment into unified safety frameworks
2. **Real-World Validation:** Deployment and testing in diverse acoustic environments (crowded streets, public transport, indoor spaces)
3. **Privacy-Preserving Architectures:** On-device processing and edge computing to minimize data transmission
4. **Low-Latency Optimization:** Lightweight models suitable for continuous background monitoring on mobile devices
5. **Standardized Evaluation Benchmarks:** Development of shared datasets and evaluation protocols for personal safety systems

V. CONCLUSION

This paper presented a comprehensive review of AI-based personal safety systems with a focus on Speech Emotion Recognition, Trigger Word Detection, and integrated safety applications. The literature survey and comparative analysis show a clear shift from traditional machine learning methods toward deep learning architectures such as CNN, LSTM, and hybrid CNN–LSTM models due to their improved accuracy and robustness.

Key Findings:

1. Deep learning models (CNN, LSTM, hybrid architectures) consistently outperform traditional ML methods, with accuracy improvements of 10–20 percentage points
2. MFCC and Mel-spectrogram features provide the best performance for speech-based tasks
3. Trigger word detection systems achieve higher precision than SER systems, making them suitable for low-false-alarm applications
4. Integrated safety systems combining multiple detection modalities show promise but remain underexplored

The analysis also reveals that most existing systems address safety components individually, highlighting a research gap in fully integrated, context-aware personal safety solutions. Overall, this review provides structured insights into current

technologies, performance trends, and limitations, which can guide future research toward developing more unified and automated AI-driven safety systems.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Ritu Chaudhary, Shreya Saraswat, Siddhant Chaturvedi, and Prof. Prajakta Naregalkar, "Speech Emotion Recognition Using Neural Network," International Research Journal of Engineering and Technology (IRJET), vol. 7, no. 5, pp. 4050–4053, May 2020.

[2] Seme Sarker, Khadija Akter, and Nursadul Mamun, "A Text Independent Speech Emotion Recognition Based on Convolutional Neural Network," International Journal of Speech Technology, vol. 26, pp. 1–13, 2023.

[3] Sheetal R. Waghmare and Prakash S. Nalbalwar, "Application of Probabilistic Neural Network for Speech Emotion Recognition," International Journal of Innovative Science and Research Technology, vol. 9, no. 2, pp. 198–202, 2024.

[4] Qianhe Ouyang, "Speech Emotion Detection Based on MFCC and CNN-LSTM Architecture," International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 13, no. 1, pp. 1–6, 2024.

[5] Shreya Jena, Sneha Basak, Himanshi Agrawal, Bunny Saini, Shilpa Gite, Ketan Kotecha, and Sultan Alfarhood, "Developing a Negative Speech Emotion Recognition Model for Safety Systems Using Deep Learning," Frontiers in Human Dynamics, 2025.

[6] Eduard Franti, Ioan Ispas, Voichita Dragomir, Monica Dascalu, Elteto Zoltan, and Ioan Cristian Stoica, "Voice Based Emotion Recognition with Convolutional Neural Networks for Companion Robots," International Conference on Automation, Quality and Testing, Robotics (AQTR), 2017.

[7] Zheng Zeng and Liang Xiao, "Effective Combination of DenseNet and BiLSTM for Keyword Spotting," 2018 IEEE 27th International Conference on Computer Communication and Networks (ICCCN), Hangzhou, China, pp. 1–8, 2018.

[8] David Bonet, Guillermo Cámbara, Fernando López, Pablo Gómez, Carlos Segura, and Jordi Luque, "Speech Enhancement for Wake-Up-Word Detection in Voice Assistants," ICASSP 2021, pp. 45674571, 2021.

[9] Arnav Kundu, Mohammad Samragh, Minsik Cho, Priyanka Padmanabhan, and Devang Naik, "HEiMDaL: Highly Efficient Method for Detection and Localization of Wake-Words," in ICASSP 2023– IEEE International Conference on Acoustics, Speech and Signal Processing, 2023, pp. 1-5.

[10] Rajath Kumar, Vaishnavi Yeruva, and Sriram Ganapathy, "On Convolutional LSTM Modeling for Joint Wake-Word Detection and Text Dependent Speaker Verification," Proceedings of Interspeech, pp. 3313–3317, 2018.

[11] Kenichi Kumatani, Sankaran Panchapagesan, Minhua Wu, Minjae Kim, Nikko Ström, Gautam Tiwari, and Arindam Mandal, "Direct Modeling of Raw Audio with DNNs for Wake Word Detection," IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2017.

[12] Zehetner, M. Hagmüller, and F. Pernkopf, "Wake-up-word spotting for mobile systems," in 22nd European Signal Processing Conference (EUSIPCO), Lisbon, Portugal, Sept. 1–5, 2014, pp. 1527–1531.

[13] Dr. R. Priya, Chiranchievi S, Santhosh M, Udhaya Parameshwaran M, and Tamizhbala S, "Safe Alert App," International Journal of Scientific Research in Engineering and Management (IJSREM), vol. 9, no. 3, pp. 15–22, 2025.

[14] Prof. Usha Kumari V, K Karthik, K Chetan, Chaithra K, and Bharath S, "SOS: Bridging the Gap Between Threat and Safety for Women Using Deep Learning Models," International Journal of Engineering Research & Technology (IJERT), vol. 13, no. 2, pp. 56–63, 2024.

[15] Sneha Dole and Dr. Santosh Jagtap, "WE SAFE App –A Women's Safety Application Using AI," International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), 2025.

[16] Zhengwei Huang, Ming Dong, Qirong Mao, and Yongzhao Zhan, "Speech Emotion Recognition Using CNN," ACM International Conference on Multimedia, pp. 801–804, 2014.

[17] Siddique Latif, Rajib Rana, Sara Khalifa, Raja Jurdak, and Julien Epps, "Direct Modelling of Speech Emotion from Raw Speech," arXiv preprint arXiv:1904.03833, 2019.

[18] Yiming Wang, Hang Lv, Daniel Povey, Lei Xie, and Sanjeev Khudanpur, "Wake Word Detection with Streaming Transformers," arXiv preprint arXiv:2103.01234, 2021.

[19] Dr. Sridhar Mandapati, Sravya Pamidi, and Sriharitha Ambati, "A Mobile Based Women Safety Application (I Safe Apps)," International Journal of Computer Science and Mobile Computing (IJCSMC), vol. 4, no. 3, pp. 252–258, 2015.

[20] Yiming Wang, Hang Lv, Daniel Povey, Lei Xie, and Sanjeev Khudanpur, "Wake Word Detection with Alignment-Free Lattice-Free MMI," Interspeech 2020, pp. 1234–1238, 2020.

[21] Liliane Momeni, Triantafyllos Afouras, Themos Stafylakis, Samuel Albanie, and Andrew Zisserman, "Seeing Wake Words: Audio-Visual Keyword Spotting," ICASSP 2020, pp. 5678–5682, 2020.