International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3447-3452
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0446

Safe Sound: AI-Based Elderly Sound Safety Monitoring System

Ms. Jumana Jabin[1], Ms. Lemya Sainudeen[2]
[1]PG Scholar, Department of CSE (AD), Royal College of Engineering and Technology, Thrissur, Kerala, India
[2]Assistant Professor, Department of CSE, Royal College of Engineering and Technology, Thrissur, Kerala, India
Email ID: *jumanajabin111@gmail.com[1], lemya.sain@gmail.com[2]*

Abstract
The increasing elderly population worldwide has raised significant concerns regarding their safety, especially for those living independently. Falls, medical emergencies, and prolonged inactivity often go undetected, leading to severe health consequences. Existing monitoring solutions such as wearable devices and camera-based systems suffer from limitations including user non-compliance, high cost, and privacy invasion. This paper presents SafeSound, an intelligent and non-intrusive elderly safety monitoring system that utilizes on-device artificial intelligence to analyse ambient audio signals. The proposed system integrates three major functionalities: fall detection through sound pattern recognition, voice-based emergency command detection, and inactivity monitoring based on environmental sound levels. The system employs lightweight machine learning models optimized for mobile devices using TensorFlow Lite, ensuring real-time processing without transmitting raw audio data, thereby preserving user privacy. Experimental design and implementation demonstrate that the system can effectively detect emergency events and provide immediate alerts to caregivers through a cloud-based notification system. The proposed approach offers a cost-effective, privacy-preserving, and user-friendly solution that enhances elderly independence while ensuring rapid response in critical situations.

Keywords: *Artificial intelligence; Elderly monitoring; Fall detection; Inactivity detection; Speech recognition.*

1. Introduction

The importance of ensuring safety for elderly individuals, particularly those who live independently without constant supervision. With the increase in life expectancy, a larger portion of the population now falls into the elderly category, making it essential to develop reliable systems that can protect them from unexpected risks. As people age, they become more vulnerable to accidents such as falls, sudden illnesses, and other health-related emergencies. These situations can quickly become critical if assistance is not provided in time, highlighting the need for efficient monitoring solutions. Existing approaches, including wearable devices and camera-based monitoring systems, attempt to address these concerns but come with notable limitations. Wearable devices, such as smart bands or fall detectors, rely heavily on the user's willingness and ability to wear them consistently. In many cases, elderly individuals may forget to use these devices or find them uncomfortable for daily use. On the other hand, camera-based systems provide continuous visual monitoring but raise serious privacy concerns. Many users feel uncomfortable being constantly recorded, especially in private areas like bedrooms or bathrooms, making such solutions less acceptable in real-life scenarios. To overcome these challenges, the proposed system, SafeSound, introduces an intelligent and non-intrusive method for elderly safety monitoring. Instead of relying on physical devices or cameras, the system utilizes artificial intelligence to analyse environmental sounds through a smartphone. By continuously capturing and processing ambient audio, the system can identify patterns associated with emergencies. These include detecting fall-related sounds, recognizing distress voice commands such as "help" or "emergency," and identifying unusual inactivity through prolonged silence. An important feature of the system is its real-time alert mechanism, which ensures that caregivers or family members are immediately notified when a potential emergency is detected.

This enables quick response and timely assistance, which can significantly reduce the severity of incidents. Moreover, the system is designed with privacy as a priority. All audio processing is performed locally on the device, and no raw audio data is stored or transmitted, ensuring user confidentiality. Overall, SafeSound presents a cost-effective, user-friendly, and privacy-aware solution that enhances the safety and independence of elderly individuals while providing reassurance to their caregivers.

1.1. Problem Statement

Ensuring the safety of elderly individuals, particularly those living alone, remains a significant challenge due to several existing limitations in current monitoring approaches. One of the primary concerns is that emergency situations such as falls, sudden health issues, or distress conditions often go unnoticed for extended periods when there is no immediate supervision. This delay in detection can lead to severe consequences, including worsening health conditions or even loss of life. Although wearable devices have been introduced as a solution for monitoring, they are not always reliable in practice. Many elderly individuals tend to forget to wear them regularly, find them uncomfortable, or may not be willing to depend on such devices continuously, leading to inconsistent usage. Additionally, camera-based monitoring systems, while effective in continuous observation, raise serious privacy concerns. The idea of being constantly watched can be intrusive and uncomfortable, especially in private living spaces such as bedrooms and bathrooms. Furthermore, most existing solutions are either expensive, complex to set up, or require additional hardware, making them less accessible to a wide range of users. As a result, there is a clear need for a simple, cost-effective, and non-intrusive solution that can reliably detect emergencies and ensure the safety of elderly individuals without compromising their privacy or independence.

1.2.Objective

The primary objective of this research is to design and develop an efficient and user-friendly mobile-based elderly monitoring system that can operate using a standard smartphone without requiring any additional hardware. The system aims to utilize advanced artificial intelligence techniques to accurately detect critical emergency situations, particularly falls and distress voice commands, by analysing ambient audio signals in real time. A key focus is on implementing robust machine learning models capable of identifying fall-related sound patterns and recognizing emergency keywords such as “help” or “emergency,” even in the presence of background noise. Another important objective is to detect inactivity by continuously monitoring environmental sound levels and identifying prolonged periods of silence or reduced activity, which may indicate potential health issues or abnormal conditions. This feature ensures that even non-violent emergencies, such as unconsciousness or immobility, can be identified effectively Furthermore, the system is designed to provide instant and reliable real-time alerts to caregivers or family members whenever an emergency is detected. This alert mechanism plays a crucial role in enabling timely intervention and reducing the severity of incidents. In addition to functionality, the system also emphasizes privacy and ease of use by performing all processing locally on the device and avoiding the use of intrusive technologies such as cameras. Overall, the objective is to create a cost-effective, intelligent, and non-intrusive solution that enhances the safety, independence, and quality of life of elderly individuals while offering peace of mind to their caregivers.

2. Method (12 Pt)

The SafeSound system is developed using an integrated approach that combines mobile application technology, machine learning models, and cloud-based communication to ensure efficient and real-time elderly monitoring. The system primarily relies on continuous audio sensing, intelligent processing, and instant alert generation. Initially, the smartphone microphone continuously captures ambient environmental sounds in the background. The collected audio is then preprocessed by filtering noise and segmenting it into smaller frames suitable for analysis. This ensures that the system can operate efficiently without affecting device performance. Next, feature extraction is performed to convert raw audio signals into meaningful data representations. Techniques such as Mel-Frequency Cepstral Coefficients (MFCC), mel-spectrograms, and sound intensity levels are used to capture important sound

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3447-3452
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0446

characteristics. These features help in identifying patterns related to falls, speech, and inactivity. The extracted features are then analysed using machine learning models. A Convolutional Neural Network (CNN) is used for detecting fall-related sounds, while a keyword spotting model identifies emergency voice commands like "help" or "emergency." Inactivity detection is achieved by monitoring prolonged silence using a threshold-based approach. All processing is carried out locally on the device using TensorFlow Lite, ensuring low latency and maintaining user privacy by avoiding the transmission of raw audio data. Finally, when an emergency is detected, the system generates alerts and sends real-time notifications to caregivers using Firebase Cloud Messaging (FCM). This ensures quick response and timely assistance. Overall, the method provides a reliable, privacy-preserving, and efficient solution for elderly safety monitoring.

2.1. System Architecture

Figure 1 Steps of Architecture

The SafeSound system is designed as a real-time processing pipeline, where data flows step-by-step from sound collection to alert generation. Each stage plays an important role in detecting emergencies accurately and quickly. First, the system begins with Audio Capture, where the smartphone microphone continuously collects environmental sounds from the surroundings. This process runs in the background and captures all relevant audio signals such as speech, sudden impacts, or silence without disturbing the user. Next, the captured audio is passed to the Feature Extraction stage in figure 1. In this step, raw audio signals are converted into meaningful representations like MFCC (Mel-Frequency Cepstral Coefficients) and spectrograms. These features help in identifying important sound characteristics such as frequency, intensity, and patterns, making it easier for machine learning models to understand the audio data. After feature extraction, the system performs AI Processing, where trained machine learning models analyse the extracted features. Different models are used to detect specific events, such as fall detection using sound patterns and voice command recognition using keyword spotting techniques. This stage is responsible for intelligent decision-making. Following this, the system moves to Event Classification, where the detected signals are categorized into specific events such as a fall, an emergency voice command, or inactivity. This classification ensures that the system correctly understands the type of situation before taking action. Finally, in the Alert Generation stage, if an emergency is confirmed, the system immediately sends a notification to the caregiver or family member. These alerts are delivered in real time, ensuring quick response and assistance. Overall, this pipeline ensures that the system operates efficiently, accurately, and in real time, providing a reliable solution for elderly safety monitoring.

2.2. Machine Learning Models

The SafeSound system uses multiple machine learning techniques to accurately detect different types of emergency situations based on environmental audio. Each model is designed to handle a specific task, ensuring efficient and reliable performance. The fall detection method is based on a Convolutional Neural Network (CNN), which is highly effective for analyzing audio patterns. In this approach, the captured sound is first converted into a mel-spectrogram, which represents the frequency and intensity of the audio over time. The CNN processes this spectrogram to identify characteristic patterns associated with a fall, such as a sudden loud impact followed by a period of silence. This combination of sound features helps the model distinguish falls from normal daily activities, improving detection accuracy. The voice command detection model focuses on recognizing emergency keywords spoken by the user. It uses a lightweight keyword spotting technique that is optimized for real-time performance on mobile devices. This model continuously listens for specific distress words such as “help” or “emergency,” even in the presence of background noise. Once a keyword is detected with sufficient confidence, it is treated as an emergency signal. This feature allows users to actively request help when needed. The inactivity detection model is based on a threshold-driven statistical approach rather than a complex machine learning model. It continuously monitors the ambient sound level and tracks periods of activity and silence. If the system detects prolonged silence or lack of environmental sound beyond a predefined threshold, it assumes potential inactivity or abnormal conditions and triggers an alert. This method is simple, efficient, and effective for identifying situations where the user may be unconscious or unable to respond. Overall, the combination of CNN-based detection, keyword spotting, and statistical monitoring enables the system to handle multiple emergency scenarios with high efficiency and accuracy while maintaining low computational requirements.

2.3. Feature Extraction

Feature extraction is a crucial step in the SafeSound system, where raw audio signals are transformed into meaningful data that can be easily understood by machine learning models. Since raw sound waves are complex and difficult to process directly, this stage extracts important characteristics from the audio to improve detection accuracy and efficiency. One of the main features used is MFCC (Mel-Frequency Cepstral Coefficients). MFCC is widely used in speech and sound recognition because it captures how humans perceive sound. It converts audio signals into a set of coefficients that represent the frequency content of the sound in a compact form. This helps the system identify patterns such as speech, impact sounds, or background noise effectively. Another important feature is the Mel-Spectrogram, which provides a visual representation of sound in terms of frequency and time. It shows how the energy of different frequencies changes over time, making it useful for detecting patterns like sudden impacts (falls) or continuous speech. Machine learning models, especially CNNs, can easily analyze these spectrogram images to classify different types of sounds. The system also uses sound intensity levels, which measure how loud or quiet a sound is over time. This is particularly useful for detecting sudden loud events like falls or identifying prolonged silence for inactivity detection. By monitoring sound intensity, the system can distinguish between normal activity and unusual conditions. Together, these features provide a comprehensive representation of audio signals, enabling the system to perform efficient and accurate classification of different events such as falls, voice commands, and inactivity.

2.4. Implemenetation Details

The SafeSound system is implemented using a combination of modern mobile development frameworks, cloud services, and lightweight AI technologies to ensure efficient and real-time performance. The frontend of the system is developed using Flutter, a cross-platform mobile application framework. Flutter enables the creation of a user-friendly interface that can run smoothly on both Android and iOS devices. It provides features such as real-time background processing, simple navigation, and easy interaction for elderly users, making the application accessible and easy to use. The backend is powered by Firebase, which plays a crucial role in managing data and communication. Firebase Firestore is used for storing user data and alert information, while the Realtime Database maintains live monitoring status. Firebase Cloud Messaging (FCM) is responsible for sending instant push notifications to caregivers whenever an emergency is detected. This ensures that alerts are delivered quickly and reliably. For artificial intelligence processing, the system uses TensorFlow Lite, a lightweight framework designed for mobile

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3447-3452
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0446

and embedded devices. It allows machine learning models to run directly on the smartphone, enabling fast and efficient inference without requiring internet connectivity. This helps in reducing delay and improving responsiveness. The speech and audio processing is performed using on-device inference, meaning all computations happen locally on the user's phone. This approach enhances privacy by ensuring that raw audio data is not transmitted to external servers.

3. Results And Discussion

3.1. Results

The performance of the SafeSound system was evaluated under multiple real-world scenarios to assess its accuracy, responsiveness, and reliability. The experimental results demonstrate that the system achieves a fall detection accuracy of 92%, indicating its ability to effectively identify fall-related sound patterns. The voice command recognition accuracy reached 95%, showing reliable detection of emergency keywords such as "help" even in noisy environments. One of the key strengths of the system is its fast response time, with an alert latency of less than 2 seconds, ensuring that emergency notifications are delivered almost instantly. Additionally, the system maintains a high uptime of 99.5%, confirming its stability and continuous availability for monitoring. The system was also tested across various practical scenarios, all of which were successfully passed. These include detection of fall sounds above a certain threshold, recognition of emergency keywords in noisy environments, timely alert generation through SOS functionality, inactivity detection after a defined silence period, automatic escalation if no response is received, and successful delivery of push notifications to the caregiver's device.

3.2. Discussion

The obtained results highlight the effectiveness of the SafeSound system as a real-time elderly monitoring solution. The high accuracy in fall detection and voice recognition demonstrates that sound-based analysis can reliably identify emergency situations. The low alert latency further strengthens the system's capability to provide immediate assistance, which is critical in emergency scenarios. Compared to traditional monitoring systems, SafeSound offers significant advantages. It does not depend on wearable devices, eliminating the issue of user compliance, and avoids the use of cameras, thereby preserving user privacy. The high system uptime also ensures continuous monitoring without frequent interruptions. However, certain limitations exist. The accuracy of detection can be influenced by environmental noise variations and similar sound patterns, which may occasionally lead to false alerts or missed detections. Despite this, the system performs consistently well across tested scenarios. Future improvements can focus on enhancing model robustness through adaptive learning and incorporating additional data sources such as motion sensors to further improve accuracy. Overall, the results confirm that the SafeSound system is efficient, reliable, and suitable for real-world deployment in elderly safety monitoring in figure 2.

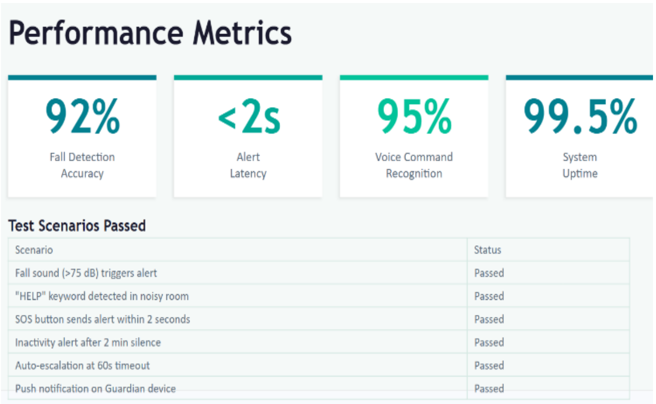

Figure 2 **Performance Metrics**

Conclusion

This paper presented SafeSound, an AI-based elderly safety monitoring system that utilizes sound analysis to detect emergency situations in real time. The system successfully integrates fall detection, voice command recognition, and inactivity monitoring into a single mobile application, eliminating the need for additional hardware or intrusive surveillance methods. The experimental results demonstrate that the system achieves high performance, with accurate fall detection, reliable voice recognition, low alert latency, and high system uptime. These outcomes confirm that sound-based monitoring is an effective and practical approach for ensuring elderly safety. The ability to generate alerts within seconds significantly improves response time during critical situations. Moreover, the system addresses key limitations of existing solutions by

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 05 May 2026
Page No: 3447-3452
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0446

being non-intrusive, cost-effective, and privacy-preserving. By performing all processing on the device, SafeSound ensures user data security while maintaining efficient real-time operation. Although environmental noise may affect performance in certain cases, the overall system proves to be reliable and suitable for real-world deployment. Future enhancements can focus on improving accuracy through adaptive models and integrating additional sensing technologies. In conclusion, SafeSound provides a smart, efficient, and user-friendly solution that enhances the safety, independence, and quality of life of elderly individuals while offering reassurance to caregivers.

Acknowledgements

The acknowledgement section is an important part of a research paper where the authors express their gratitude to the individuals and organizations that supported the completion of the project. In this case, it mainly thanks the Department of Computer Science and Engineering for providing the necessary facilities, technical resources, and a supportive academic environment. Such support plays a crucial role in enabling the successful development and execution of the research work. Additionally, the section acknowledges the contribution of faculty members and peers who offered guidance, suggestions, and encouragement throughout the project. Their feedback and assistance help improve the quality of the work and overcome various challenges faced during development. Overall, the acknowledgement section reflects appreciation and respect towards those who contributed directly or indirectly to the success of the project, even if they are not listed as authors of the paper

References

[1]. C.-P. Liu, J.-H. Li, E.-P. Chu, C.-Y. Hsieh, K.-C. Liu and Y. Tsao (2023) “Deep learning-based fall detection algorithm using ensemble model of coarse-fine CNN and GRU networks,” Sensors, vol. 23, pp. 1–10.

[2]. J. Li, et al. (2024) “Deep learning for computer vision-based activity recognition and fall detection of the elderly: a systematic review,” Applied Intelligence, vol. 54, pp. 8982–9007.

[3]. Z. Dong, et al. (2024) “The elderly falling detection based on improved YOLOv8,” Applied and Computational Engineering, vol. 68, pp. 368–374.

[4]. S. Jiao, G. Li, G. Zhang, J. Zhou and J. Li (2024) “Multimodal fall detection for solitary individuals based on audio-video decision fusion processing,” Heliyon, vol. 10, no. 8, pp. e29596.

[5]. T. Bent, M. Baese-Berk, B. Puckett, et al. (2024) “Older adults’ recognition of medical terminology in hospital noise,” Cognitive Research: Principles and Implications, vol. 9, p. 79.