Grenze International Journal of Engineering and Technology, June Issue

Vital Eyes – An AI-Powered Multimodal Fall Detection System

A Kavitha[1], Netra Srinivasan[2], Prathana Y[3] and Raja Nandhini S[4]
[1-4]Easwari Engineering College/Electronics and Communication Engineering, Chennai, India
Email: kavivenkat99@gmail.com, netra.sriala20@gmail.com, yprathanayuvaraj@gmail.com, nandhinisr56@gmail.com

***Abstract*— India's elderly population has reached 140 million, and supporting independent aging is significant. Falls have become very common among the senior citizens, leading to undesirable situations and accidents. Existing devices provide limited support and lack immediate response, leading to imprecise decision making. Our system aims to resolve this by providing AI-powered emergency monitoring to mitigate after-effects of elderly falls. This system continuously monitors patient vitals, posture, sound classification and gesture detection and effectively makes a decision cumulating all the parameters, providing patient's condition and immediately alerts when fall occurs. Through the MediaPipe algorithm, the machine learns abnormal gestures. It has three main significances: biomedical data analysis, audio detection of normal or abnormal sounds, and precise body monitoring using MediaPipe, which continuously provides the exact position. Based on these three factors, we can obtain a clear decision and determine the exact state of the patient. Whenever there is an abnormality, the system triggers an alert to the caretaker or the pre-registered responsive person. This enhances safety and strengthens the digital healthcare system. As the population rises, the care should also increase. As a future scope, it can incorporate real sensing and optimization for low-power hardware and enable large-scale deployment across homes and institutions. The ultimate aim is to create smart safety system and efficient fall detection to empower elderly citizens to live independently with confidence, dignity and security.**

***Index Terms*— AI-enforced Healthcare, Elderly Safety Monitoring, Machine Algorithm Learning, MediaPipe and Audio Classification, Biomedical Signal Analysis, Digital Healthcare Innovation.**

I. INTRODUCTION

The world has an increasing population, whereas in India it is over 140 million and includes more senior citizens. As technology develops, it also increases risks towards healthcare. As the population increases, the healthcare technology to take care of the population also increases. Elders are the backbone for everyone, but taking care of them has turned insignificant these modern days. The possibilities of their fall have increased, leading to undesirable events. Elders, therefore face great challenge to live independently. Current methodologies, such as [1], [2], [6] and [10], focus on fall detection using isolated approaches – IoT-based sensing, MediaPipe, or keypoint extraction – and often consider fall detection and patient monitoring as separate tasks, which might limit accuracy and raise false alarms, creating unwanted panic for families. This demands a more accurate system and an advancement of healthcare.

Grenze ID: 01.GIJET.12.2.55_22

VitalEyes is enabled with a continuous monitoring system for the patient with abnormality detection that integrates the following parameters, such as biomedical analysis of data, tracking of motion, and algorithm-based audio classification to recognize body posture and identify and analyze abnormalities, which is provided for 24×7 real-time continuous monitoring. Whenever there is an abnormality, the trigger will be enabled after considering all the factors. This system exhibits accurate decision-making and reduces fall detection, which recognizes the fall of the patient during abnormal conditions. This system is specified for the improvement of healthcare for elder protection and their freedom to live an independent life. This vital eye is trustworthy and helps the elder to live safely and with dignity.

II. METHODOLOGY

The VitalEyes is a multi-modal artificial intelligence-powered solution for enhancing the safety of elderly people to live an independent life. It incorporates multiple sensor inputs, which include biomedical, audio, and visual data with the help of gesture identification. By classifying abnormal and normal cases, this system enables and implements decisions to identify the condition of the patient. This system detects elderly fall by integrating three inputs, such as audio-based signals, sensing of biomedical data, and visual or gesture identification from the continuous real-time data. This multi-modal detection exhibits accuracy with improvement in reliability and efficiency shown in Figure 1,2 and 3.
The multimodal health monitoring system includes,

- Biomedical data collection of Temperature, SPO_2 , Blood Pressure and ECG.
- IMU-based gait & impact analysis from six body locations (right/left feet, shins, thighs).
- AI-driven audio + gesture detectors

The overall fusion in a decision logic engine issue local (buzzer) and remote (SMS/Email via Bluetooth gateway) alerts when appropriate.

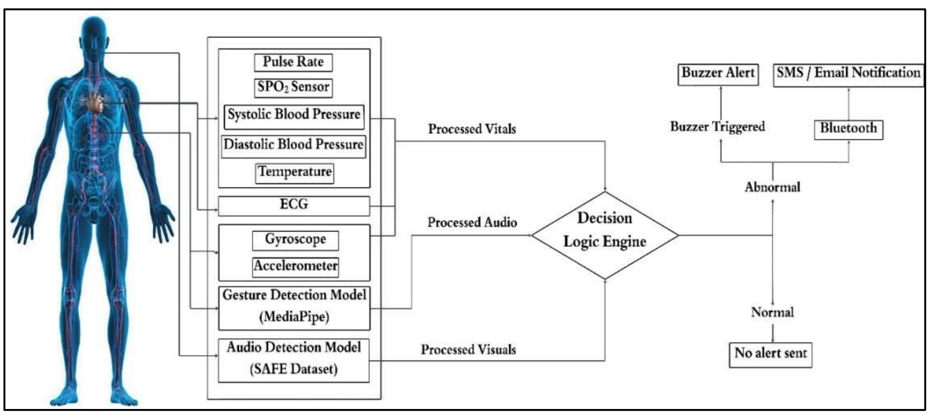

Figure 1. Block Diagram of Multimodal Sensor & Decision Logic

A. Virtual Vitals Monitoring

Heart Rate

Heart Rate (HR) is derived from inter-beat intervals (RR):

$$\mathrm{HR} = \frac{60}{\mathrm{RR\_interval}} \qquad (1)$$

Heart rate equation is referred from [3]. RR_interval is measured in seconds. This formula converts the cardiac cycle period into beats per minute (bpm).

Oxygen Saturation (SpO_2)

The formula that is derived for calculating the threshold value is

$$SpO_2 \approx f\left(\frac{AC_{red}/DC_{red}}{AC_{IR}/DC_{IR}}\right) \qquad (2)$$

This equation is taken from the reference [7],where the ratios exibits the pulsatile red wavelength and baseline IR wavelength.

Blood Pressure (Cuffless Estimation)

The system is based on surrogate and pulse transit time, it expresses the Blood pressure threshold equation,

$$\mathrm{BP} = \alpha + \beta(\mathrm{PTT}^{-1}) + \gamma \cdot \mathrm{HR} + \varepsilon \qquad (3)$$

This formula is referred from the [8]. where alpha, beta, gamma are the calibrated coefficient from model fitting.

Body Temperature

The threshold range that indicated the range for temperature calculation is,

$$36.1°\mathrm{C} \leq \mathrm{T} \leq 37.2°\mathrm{C} \qquad (4)$$

This is the general range for the temperature and is taken from [5].

Electrocardiogram (ECG)Signal Processing

The PTB-XL dataset is used in the ECG calculation as a dataset which is stored with digital signals sampled as uniform with regards to time

$$X = \left[x_l[n]\right], n = 1,2, \dots, N; \;\; l = 1,2, \dots, L \qquad (5)$$

The classification C(X) is expressed as

$$C(X) = \begin{cases} Abnormal \;\; if \; max_{n,l} \; x_l[n] > \; T_{high} \; or \; min_{n,l} \; x_l[n] < \; T_{low} \\ Normal \qquad\qquad\qquad\qquad\qquad\qquad\qquad otherwise \end{cases} \qquad (6)$$

This rule exhibits and distinguish the flag recording with electrical atypical activity and defines the normality and abnormality.

B. IMU (Gait & Impact) Processing — Conversion, Filtering & Formula [3][4]

Sensor Conversion

$$a = \frac{\mathrm{raw}}{16384} \times 9.81 \qquad (7)$$

This equation for physical acceleration is referred from [4]. Raw indicated an int16 sensor which calculates for ±2g range with the value of 16384 counts/g.

Immobility Check (Post-Impact)

The post-impact calculation mean deviation ($\Delta \mathrm{a}_{\mathrm{post}}$) as the gravity generated over the N samples with the windows (Window Time = Duration of the immobility):

$$\Delta \mathrm{a}_{\mathrm{post}} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left| |\mathrm{a}(\mathrm{t_i})| - \mathrm{g} \right| \qquad (8)$$

Formula for immobility check is taken from reference [4]. $|\mathrm{a}(\mathrm{t_i})|$ denotes the magnitude of acceleration at time $\mathrm{t_i}$ and g denotes the standard acceleration due to gravity. If $\Delta \mathrm{a}_{\mathrm{post}} < \delta_{imm}$ (with $\delta_{imm} = 2.0 \; m/s^2$ is the immobility duration), the sensor is labelled as immobile.

Multi-Sensor Fusion

This process uses a time-window tolerance T_{tol} (T_{tol} =0.2s): merge impacts whose times differ by $\leq T_{tol}$. The fused fall candidate uses average time:

$$\mathrm{t_{fall}} = \frac{1}{\mathrm{M}} \sum_{m=1}^{M} \mathrm{t}_{\mathrm{impact}}^{m} \qquad (9)$$

Formula for fall time calculation is referred from reference [13]. Averaging mitigates timestamp jitter between sensors and provides a single fall timestamp for alerts, referred from [7][8].

C. AI Modules (Audio & Gesture) — Model Description & Integration

Audio Detection

The audio module is implemented using MFCC-based feature extraction followed by a CNN classifier. Each audio clip is first resampled at 16 KHz with fixed duration 3 s. If the clip length L is shorter than the target window, it is zero-padded; if longer, it is truncated, referred from [13]:

$$x'[n] = \begin{cases} x[n] & 0 \le n \le L \le sr.d \\ pad(x[n]) & L < sr.d \end{cases} \qquad (10)$$

Formula for audio pre-processing is taken from reference [9]. Here x[n] is the original audio signal and x'[n] is the pre-processed audio signal. From this signal x'[n], Mel-Frequency Cepstral Coefficients (MFCCs) are extracted.

$$f = \frac{1}{T}\sum_{t=1}^{T} MFCC(x'[t]) \qquad (11)$$

Formulas [8] and [10] for MFCC and feature vector computations is taken from reference [12]. Here f represents the feature vector t represents the time frame within audio signal, $\frac{1}{T}$ is the average of MFCC vectors to get a single feature representation per audio clip.

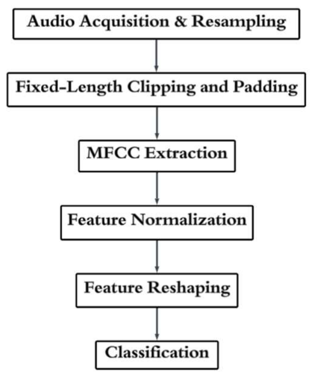

Figure 2. Audio Detection Workflow

Gesture Detection (via MediaPipe)

Skeletal key-points K_t are extracted frame-by-frame using a pose-estimation backbone [5]. A temporal classifier C_{gest} evaluates sequences of skeletons over window $t - T{:}\,t$

$$P_{gest}(t) = C_{gest}(X_{t-T:t}) \qquad (12)$$

Formula for representation of gesture data feature is taken from reference [6]. $P_{gest}(t)$ is the feature representation of gesture data at time t.

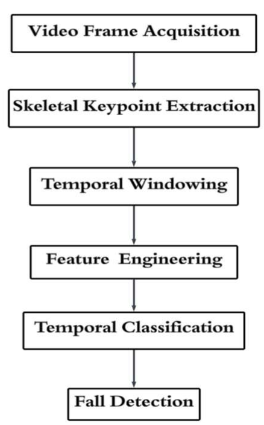

Figure 3. Gesture Detection Workflow

D. Final fall decision combining Vitals, Multimodal AI and Alert System

To enhance the robustness of fall detection, a multimodal decision fusion frame work implemented. Gesture recognition is assigned the highest priority, referred from [7,15], vital sign deviations are integrated as a secondary modality is taken from the reference [6,7,12]. This workflow improves the efficiency of the system, thereby ensuring reliability.

E. Decision Logic Formulation

The decision logic of VitalEyes can be formulated as:

$$Alert = P \lor S \lor (A \land (P \lor S)) \tag{13}$$

where P indicates fall detected on posture and gesture,
S corresponds abnormalities in sensor data and
A represent to audio signal detection

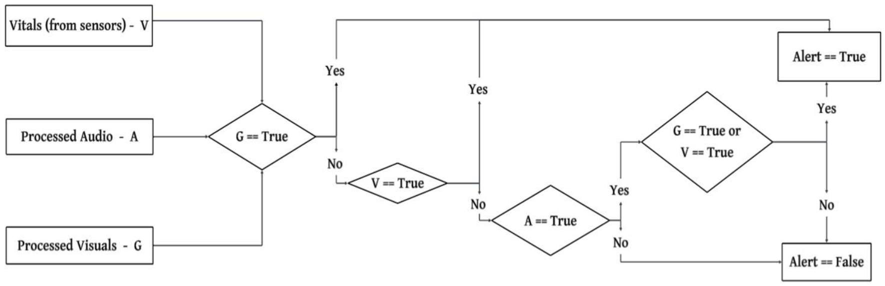

Figure 4. Decision Logic Execution

III. Results and Discussion

The performance of the implemented fall detection framework is analyzed.

A. Vitals- Testing

This involved processing each parameter and combining the outcomes to provide an overall analysis of patient based on vitals.

Vital Sensors Simulation

The integrated output with the sensor conditions of blood pressure, temperature, SpO_2, and with the ECG signal output. Based upon the pre-processing conditions, these sensor outputs are combined to get the overall decision output. Since vital outputs are calculated separately and then the ECG output is calculated, together integrate the ECG and vital outputs together, which gives a precise biomedical output of normal or abnormal based on the patient's condition with true or false alerts. And ECG is prioritized as high weightage and the vitals for overall health analysis and combined alert system for vitals as shown in Figure 4 and 5.

```
File: 01000_hr | G: 1.00 | E: 0.0 | V: 0.62 | Score: 0.59 | Fall: True
File: 01001_hr | G: 1.00 | E: 0.0 | V: 0.62 | Score: 0.59 | Fall: True
File: 01002_hr | G: 1.00 | E: 0.0 | V: 0.62 | Score: 0.59 | Fall: True
File: 01003_hr | G: 1.00 | E: 0.0 | V: 0.88 | Score: 0.66 | Fall: True
File: 01004_hr | G: 1.00 | E: 0.0 | V: 0.88 | Score: 0.66 | Fall: True
File: 01005_hr | G: 1.00 | E: 0.0 | V: 1.00 | Score: 0.70 | Fall: True
File: 01006_hr | G: 1.00 | E: 0.2 | V: 0.75 | Score: 0.69 | Fall: True
File: 01007_hr | G: 1.00 | E: 0.0 | V: 0.75 | Score: 0.62 | Fall: True
File: 01008_hr | G: 1.00 | E: 0.0 | V: 0.62 | Score: 0.59 | Fall: True
File: 01009_hr | G: 0.00 | E: 0.0 | V: 0.88 | Score: 0.26 | Fall: False
File: 01010_hr | G: 1.00 | E: 0.0 | V: 0.88 | Score: 0.66 | Fall: True
File: 01011_hr | G: 1.00 | E: 0.0 | V: 0.88 | Score: 0.66 | Fall: True
File: 01012_hr | G: 0.00 | E: 0.0 | V: 0.88 | Score: 0.26 | Fall: False
```

Figure 5. Cumulated Vitals and Postural State Outcome

B. Audio – Testing

The audio phase is used to classify the sound signal from the pressure to electrical signal, and the electrical signal is converted through ADC to DAC for digital amplification graph with the help of Fourier transform analysis. And we use middle frequency to process it as a human audible sound, which helps to classify the overall output of audio as normal or abnormal.

MEL Frequency Processing

The Mel spectrogram is scaled to match human auditory perception. Different energy patterns indicated different sound frequencies (normal and abnormal). This made audio testing more accurate to detect health-related anomalies as shown in Figure 6.

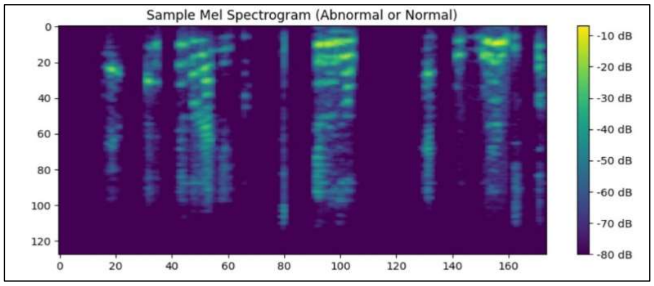

Figure 6. Mel Frequency of sample audio

C. Gesture Detection

Evaluation of body posture and movement for accurate fall identification was performed. The system classified normal and abnormal gestures.

Pose Keypoint Extraction

In this stage, MediaPipe algorithm was deployed to obtain 33 keypoints (x, y, z, visibility) from each video frame. On classifying the missing detections as NaN, videos of fall and normal postures were processed. The extracted features were stored in frames_keypoints.csv, while FPS information was cached separately. This dataset forms the raw input for later processing as shown in Figure 7.

```
              precision    recall  f1-score   support

           0       1.00      0.96      0.98      1420
           1       0.95      1.00      0.97      1092

    accuracy                           0.98      2512
   macro avg       0.98      0.98      0.98      2512
weighted avg       0.98      0.98      0.98      2512

Confusion:
 [[1366   54]
 [   2 1090]]
```

Figure 7. Performance Benchmark

Model Training

A Random Forest Classifier (RFC) was trained on the extracted window-level features. The model achieved good classification accuracy, with confusion matrix and classification report highlighting fall vs non-fall separation. The feature list and trained model were saved for inference.

D. Real-Time Inference and Alert

The model trained (incorporating vitals, audio and gesture parameters) was deployed for real-time fall detection using live and recorded videos. The obtained video will be exhibited as each frame, and each frame deciding window technique is used to classify the normality and abnormality based upon the threshold condition predefined, and if in case of any abnormality, the sound will be triggered and also the message on the screen of the device will be popped up. And this is verified with the camera position and also with varied conditions, and integrity with reliability. The overall framework is subsequently deducted with a minimal false alarm deducted as shown in Figure 8.

```
--- Decision Report ---
Audio:   pred=1, prob=0.68
Gesture: pred=0, prob=0.28
Vitals:  pred=1, prob=1.00
=> FINAL FALL DECISION: 1 (score=0.63)
```

Figure 8. Decision Output

The Blaze Pose Deep Learning System, with the help of the Media Pipe algorithm for the project VoLTE Reye, showed 98% accuracy reduction on pointing out the pose exactly. The existing fall detection using Media Pipe implementation has attained an accuracy of 90%, where the VoLTE Reye exhibits 98% with 8% improved accuracy on combining the parameters of all health-related conditions as shown in Figure 9.

Figure 9. Fall Detection Result

IV. CONCLUSIONS

This multimodal emergency call direction framework combines, along with the real-time monitoring and continuous health condition analysis with the help of vital and ECG analysis and AI-based audio detection and gesture detection through the frames of the visuals consideration, as this work focuses on the safety of the patient with the threshold-based value and continuous monitoring through the sensor integration and physiological fusion. With the help of deep learning models for analyzing audio and gestures, the system enables accurate fall detection and reduces false alarms, which also generates alerts to the caretaker. This quick intervention prevents injury to people who need special care, as it is an elder-specific design focusing on effective processing for detecting falls and prioritizing their privacy by securely transmitting alerts through both remote channels and locally. Vital Eyes aims to enhance the quality of elders' lives in a responsible manner socially by connecting state-of-the-art deep learning and artificial intelligence with practical healthcare focus.

LIMITATIONS AND FUTURE SCOPE

Vital Eyes is a real-time continuous monitoring system with certain limitations, which can be enhanced with future extensions. First, occasionally inaccurate sensor readings occur depending on the body position of the individuals and their movements, which can be resolved with adaptive thresholds on body movements and could be incorporated as an enhancement in future work to adjust sensor data variations. Second, the ESP32 is a microcontroller with RAM and ROM-based memory, and it is limited in memory space. This limitation can be addressed by deploying model training with memory-efficient resources using the TinyML technique instead of huge datasets. Tiny ML enables lightweight deployment of model training, which creates space and reduces the consumption of large memory. Third, it is a prototype with the current implementation exhibited using datasets, whereas real-time patient data and system integration might be complicated, which can be addressed through dynamic handling of data and allocation using weights.

REFERENCES

[1] Ahmed, N., Mahfuz, H., Haque, M.S. and Taheri Nejad, N. (2022) 'Stress and Fall Detection via IoT', IEEE Sensors Journal.
[2] Aly, T., El-Din, N.G. and Ahmad, I.H. (2022) 'Remote Patient Health Monitoring in IoT', Journal of Medical Systems.
[3] Altnas, E., Etemadi, M.A., Ataei, S., AlCheikh, G. and Jorswieck, E. (2022) 'Monitoring Cardiopulmonary Function: wearable sensors and algorithms', US Patent 11395595 B2, Jul 26.
[4] Banerjee, S., Ahmad, I.Z. and Mukherjee, N. (2022) 'Low-G Impact Fall Detection with Tri-Axial Accelerometer', IEEE Sensors Journal.
[5] Choudhury, D., Ahmed, S. and Rizwan, F. (2023) 'Wearable Health: Multimodal Sensing', Computers & Electrical Engineering.
[6] Chouhan, W., Sahore, N. and Agrawal, A. (2023) 'MediaPipe-based Real-time Fall Detection', Proceedings of ICECE.
[7] Eldardiry, H.M.A., Rubin, J., Abreu, R., Ahern, S.P., Bobrow, D.G., Garcia, D., Du, H. and Pattekar, A. (2017) 'Interactive remote patient monitoring and condition management intervention system', US Patent US20170049374A1.
[8] Kavitha, A. and Swaminathan, J.N. (2023) 'Design of IoT Based Human Bond Communication in Smart Foam
[9] Nanocomposite Sensor Through In-Situ Polymerization Method for Female Child Safety', Journal of Electrical Engineering & Technology, April.
[10] Kavitha, A., Hemalatha, B., Abishek, K. and Harigokul, R. (2023) 'Internet of Things (IoT) based wearable device for monitoring women health', Proceedings of the 4th International Conference on Circuits, Control, Communication and Computing (I4C).

[11] Kaur, J., Handa, A. and Khokhar, K.B.S. (2022) ‘Vision Based Fall Detection System Using Keypoints and Ensemble Classifiers’, Proceedings of I4C.
[12] Kim, S. and Hwang, G.-S. (2017) ‘Device for calculating systolic blood pressure using pulse transit time and method therefor’, PCT Patent WO2017010832A1.
[13] Kumar, A., Khambete, N. and Naskar, M.K. (2023) ‘Mobile Agent IoT Healthcare’, Proceedings of DEXA.
[14] Kumar, R. and Sendil, K. (2023) ‘Smart Healthcare IoT Framework’, Sensors (Basel).
[15] LeBoeuf, S.F. and Pattekar, A. (2020) ‘Systems for Cleaner Physiological Signals’, US Patent 10716526 B2, Jul 21.
[16] LeBoeuf, S.F., Rykstad, R. and Manoni, R.J. (2022) ‘Cleaner Physiological Signals with Motion Elimination Techniques’, US Patent 11324407 B2, Apr 12.