

SafeHer AI

Prof. Poonam Kapse, Prof. Vishwayogita A. Savalkar, Sneha Mestri*, Sahil Kanse, Shivam Mohite,
Saurav Rai, Rishi Chilveri
Department of AIML

Bharati Vidyapeeth Deemed to be University, Department of Engineering & Technology, Kharghar,
Navi Mumbai, Maharashtra, India

*Corresponding Author – Email ID: smestri842@gmail.com

Abstract

In particular, the lack of immediate human assistance is a major reason why women's safety continues to be a problem in society at large. Traditional safety apps mostly require the user to manually interact with them, so they are not helpful at all in the case of a sudden panic or physical distress. This article presents an AI-powered smart wristband system that can detect unsafe situations in real time and respond to emergencies automatically. The system under discussion uses wearable sensors such as an accelerometer-gyroscope module, a heart rate sensor, a microcontroller, a base embedded platform, and a smartphone interface. Wearable device motion data are used in this research work to locate abnormal or panic-like movement patterns by means of a Temporal Convolutional Network (TCN), whereas physiological stress can be detected on the basis of personalized heart rate analysis committed to individual baseline variations. Furthermore, through GPS-based risk zone analysis and time-dependent rules, the system's sensitivity could be dynamically changed, and hence achievement of contextual awareness. A multi-modal decision framework, which is based on user confirmation, is employed in order to minimize false alarm generation from bio signal interpretation. This mechanism is followed by the incorporation of physical features and personal health features coupled with motion that are then subjected to an analysis through a multi-modal decision framework, which results in the detection of the event and alarm generation. Once the system has figured out that the threat is real, it will not hesitate to send out an SOS alert as well as allow someone to track the live location of the person continuously through a cloud platform. This proposed model is a personalized, multi-sensor fusion and low-latency decision, approach that has been compared to a reliable and scalable solution for women.

Keywords – Women's Safety, Smart Wearable Device, Temporal Convolutional Network, Stress and Panic Detection, Sensor Fusion, Context-Aware Emergency System.

1. Introduction

The Internet of Things (IoT) has turned out to be a leading computing paradigm that basically makes it possible for everyday physical objects to get connected with digital systems via the internet. Typically, such objects come with several components, including sensors, processors, and communication technologies. Some very rudimentary types of IoT were around in the 1980s when a handful of trial setups demonstrated the possibility of linking physical devices to computer networks for their remote monitoring and control. Nevertheless, the overall idea was not really identified until 1999 when Kevin Ashton coined the term Internet of Things and pointed out that embedded identification technologies play a bridging role between the physical and the digital worlds [1]. In the early 2000s, advancements in features such as low-power microcontrollers, wireless communication protocols, and internet infrastructure were so rapid that, finally, IoT was no longer limited to being experimental demonstrations, but at the same time, leveraging the various sectors. The massive popularization of cloud computing and the opening up of IPv6 addressing capabilities have led to an extraordinary expansion of IoT ecosystems, which are now capable of connecting billions of devices worldwide [2]. Currently, IoT is serving as the core technology for smart environments such as remote health monitoring, factory automation, smart cities, and personal security systems, therefore facilitating continuous data collection, advanced analytics, and autonomous decision-making [3]. Artificial Intelligence (AIoT) Internet of Things (IoT) systems are essentially the fusion of Artificial Intelligence (AI) and IoT. It describes a situation where the information gathered from a series of devices connected to the network is analyzed and interpreted by AI algorithms. Generally, typical IoT systems are mostly designed for sensing and communication operations [4]. Nevertheless, by integrating AI, the systems can perform complex analytics, detect patterns, and even decide based on predictions, thus enhancing the system's autonomy and trustworthiness. Machine learning (ML) and deep learning (DL) methods

can process large volumes of different sensor data, thereby facilitating various applications such as heightened accuracy in activity recognition, detection of unusual events, and prediction of behaviors in real time[5]. Moreover, AI embedded in the edge of IoT systems can provide nearly real-time responses, reduce the dependence on cloud servers, and improve data security, thus making AIoT a perfect fit for high safety requirements and wearable applications [6].

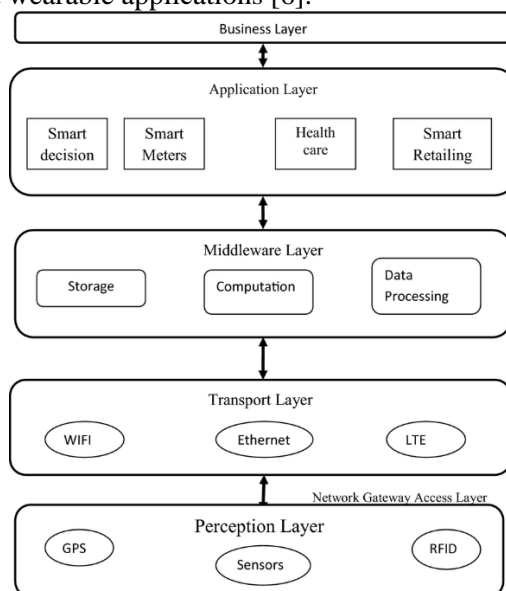


Figure 1: Layer of IoT Architecture

IoT Layer Architecture:

1. **IoT Devices / Sensors:** The layer is comprised of the wearable sensors that are embedded in a smart wristband to collect the physical and physiological data. The MPU6050 accelerometer and gyroscope sensor module is used to continuously collect the user movement and gesture patterns. The MAX30102 sensor is used to monitor physiological signals that are associated with stress and panic attacks and these sensors provide the raw data, which is used as the primary input to be intelligently analyzed.
2. **Edge Layer (AI at Edge):** The edge layer has been implemented by using the ESP32 microcontroller, which comes into direct contact with the wearable sensors. The real-time acquisition and processing of the data, as well as the inference, can be done using the microcontroller. The motion data collected is processed using the Temporal Convolutional Network (TCN), which helps in the identification of the activities. The heart rate data collected from the sensors can be processed using the personalized baseline logic for abnormal stress conditions. To achieve minimum latency and instant response in the emergency situations, edge processing has been utilized.
3. **Communication Layer:** This layer enables the communication between different parts of the system through the wireless communication and the data received by the ESP32 is transmitted to the phone via Bluetooth. The phone is then connected to the cloud and the communication between wearable, app, and cloud is facilitated by multi-layer communication system.
4. **Cloud Layer:** Firebase is the cloud layer, which combines the capabilities of data storage and live tracking. The cloud primarily stores the emergency alerts and the GPS coordinates that received from the smartphone. The cloud maintains the live tracking of the location shared with the emergency contacts. The live tracking is continued until the emergency alert is turned off.
5. **AI / ML Layer:** The role of the AI/ML part is to serve as the brain of the system, that combines deep learning technology in the motion analysis with a temporal convolutional network, and stress detection through a rule-based or lighter model. It also utilizes personalized AI to make the accuracy better by tailoring the thresholds to each user. The system also uses multi-modal sensor

fusion to integrate motion risk, stress risk, and contextual risk and combine different streams of data to measure the potential risk for smarter decisions.

- 6. Application Layer:** The app is the platform where users interact with their alerts on their phones. The app has GPS, a feature to send emergency alerts, confirmation, and live monitoring. Once the system has confirmed that the danger is real, it sends an SOS to the cloud. At the same time, it sends the updated location of the person to his or her pre-designated emergency contacts. This helps ensure a timely response to keep the person safe.

2. Literature Survey

Citations	Paper Title	Paper Type/ Source of Paper	Technology used	Dataset used	Result	Limitation
[1]	Danger Det: A mobile application based danger detection platform for women and children using deep learning	Journal Published on Science Direct	1.MobileNetV 2. InceptionV3 (with Noise Reduction)	Audio Dataset (~2,000 samples, 3 classes: normal, woman in danger, child in danger)	Accuracy : 1. 93.8% (offline) 2. 95% (online)	1. Offline version requires more storage and slows device 2. Online version uses mobile data.
[2]	Real-time active learning method for audio-based anomalous event detection	Journal Published on Frontiers in Computer Science	1. Random Forest 2. Multilayer Feedforward Neural Network	AEDAF Audio Dataset (21,920 samples, 10 abnormal events).	1. 99.26% (AELM) 2. 95% (DL model)	1. Computationally intensive. 2. May require frequent retraining
[3]	From Vision to Sound: It is Advancing the Audio Anomaly Detection with Vision based Algorithms	Journal Published on Research Gate [2025]	1. VAD Algorithms 2. PaDiM (Patch Distribution Modeling) 3. CFA (Clustering Feature)	MIMII & EnvMix Datasets (~26k samples, normal and anomalous machine/environment sounds).	1. ROC up to 90.6% (sample) 2. 82.9% (temporal) 3.76.4% (spectrogram)	Visual modality integration may increase complexity and power usage.
[4]	Unified AI for Accurate Audio Anomaly Detection	Journal Published on Research Gate [2025]	1. Random Forest 2. SVM Model 3. CNN Model	Unified AI for Audio Anomaly Detection (~1.5k speech/anomaly samples)	1. 96.8% accuracy. 2. 96.2% precision 3. 97.1% recall	Ensemble models may be resource-heavy
[5]	Fundamental Survey on Neuromorphic Based Audio Classification	Journal Published on Research Gate [2025]	Spiking Neural Network	Neuromorphic Audio Classification Survey (~2k event-based and spiking audio samples)	1.92% detection accuracy 2. Precision = 90% 3. Recall = 93%.	May have lower accuracy compared to conventional deep learning.
[6]	Quick detection and classification of hazardous urban sounds	Journal Published on Research Gate [2025]	1. CNN Model 2. RNN Model	Dangerous Urban Sounds Dataset (~2.5k audio samples)	1. Accuracy ranged from 79% to 93% highest for siren 93.06% and fire 92.03% 2. AUC ROC up to 0.95	Might have lower accuracy during the detection and classification sound.

System Architecture and Hardware:

Component	Stress Detection Module	Motion Gesture Module
Model Architecture	LSTM Autoencoder	Temporal Convolutional Network (TCN)
Primary Dataset	WESAD (Wearable Stress and Affect Detection)	WISDM (Wireless Sensor Data Management)
Input Data Types	Physiological signals (likely BVP, EDA, ECG, etc.)	Tri-axial Accelerometer signals (<i>accel_x</i> , <i>accel_y</i> , <i>accel_z</i>)
Detection Method	Reconstruction Error (Unsupervised/Anomaly Detection)	Categorical Classification (Supervised Learning)
Key Output Classes	Non-Stress vs. Stress	Walking, Jogging, Upstairs, Downstairs, Sitting, Standing
Performance Metrics	ROC-AUC (0.74), Precision, Recall, F1-score	Accuracy (approx. 95%), Per-class Accuracy, Loss

MPU6050 (Accelerometer + Gyro): Connected to an ESP32 for motion and gesture sensing.

Heart Rate Sensor (MAX30102): Connected to an ESP32 for detecting stress or panic levels.

ESP32 Microcontroller: It acts as a central hub for the local processing and communication via Bluetooth.

Smartphone: The system uses communication techniques with the Bluetooth with ESP32, and offers GPS, internet, and an interface to communicate.

2. Methodology:

A smart wristband, with aid of AI, detects dangerous moves through smart motion analysis, monitors physiological stress, and tracks your location in real-time, taking into consideration the surrounding context to assess the risk. The wearable device uses the deep learning with rule-based logic to detect the potential threats. Conventional safety devices require pressing an SOS button to activate them. This new system uses the wearer’s AI and multi-modal sensing and confirmation to minimize false alarms and ensure a quick response to real emergencies.

Hardware–Software Integration Architecture:

It uses hardware and software that are integrated in such a manner that the MPU6050 module controls the motion data collected from the accelerometer and gyroscope and is directly linked to the ESP32 microcontroller for the detection of gestures and movements. The MAX30102 heart rate sensor is linked to the ESP32 microcontroller for the detection of physiological stress indicators. The ESP32 performs the local calculations, while the smartphone is used to handle the user interface through Bluetooth Low Energy. The GPS and internet access of the smartphone provide the interface to the user. Firebase is used to provide the emergency alerts and the real-time tracking to aid the decision-making process.

Sensor Data Acquisition:

By using this approach, movement and body signals were possible, and MPU6050 were used to get information regarding how the body is moving, including its acceleration and rotation. On the other hand, the MAX30102 sensor offers data on the body’s heart rate changes, which can indicate abnormal stress in the body. This data is processed by the ESP32 microcontroller after pre-processing which is an important process because it sets the stage for effective AI inference.

Motion Analysis Using Temporal Convolutional Network (TCN):

To identify motions, we begin with cutting the raw data from the accelerometer into smaller equal-sized segments and thus creating a structured time series input. These inputs are then passed through a Temporal Convolutional Network (TCN), a type of neural network that performs particularly well when working with sequences. The TCN not only identifies normal activities such as walking and jogging but also recognizes panic motions that indicate distress. We are using TCNs because we observe that they provide not only low latency but also steady gradients and performance compared to the other two prominent models: CNN and LSTM.

Physiological Stress and Panic Detection:

Besides analyzing movement, we also want to look at heart rate over small windows to detect signs of panic or sudden stress. Rather than relying on fixed, generic thresholds, we want to weigh how heart rate is changing relative to an individual's own normal state. A simple model for classifying stress levels—or a rules-based decision process—can then determine if the body state is normal or if it is a state of panic. By decoupling movement analysis from stress detection, we can better manage multiple sensing modalities to increase robustness and accuracy.

User Personalization Mechanism:

A major innovation of the proposed system is its ability to tailor itself to the individual user while using the device primarily; it also captures and saves normal levels of resting movement and heart rate of individual user. The saved levels are then used to update the levels of the detection thresholds. This adaptive approach, which accounts for individual variations in physical characteristics and differences in working modes, ensures that all users experience a consistent level of service without interruptions. Including user-specific adjustment, the system will be able to directly correspond to person-to-person differences, which is one of the most common drawbacks of widely available safety monitoring systems.

Context-Aware Risk Zone Analysis

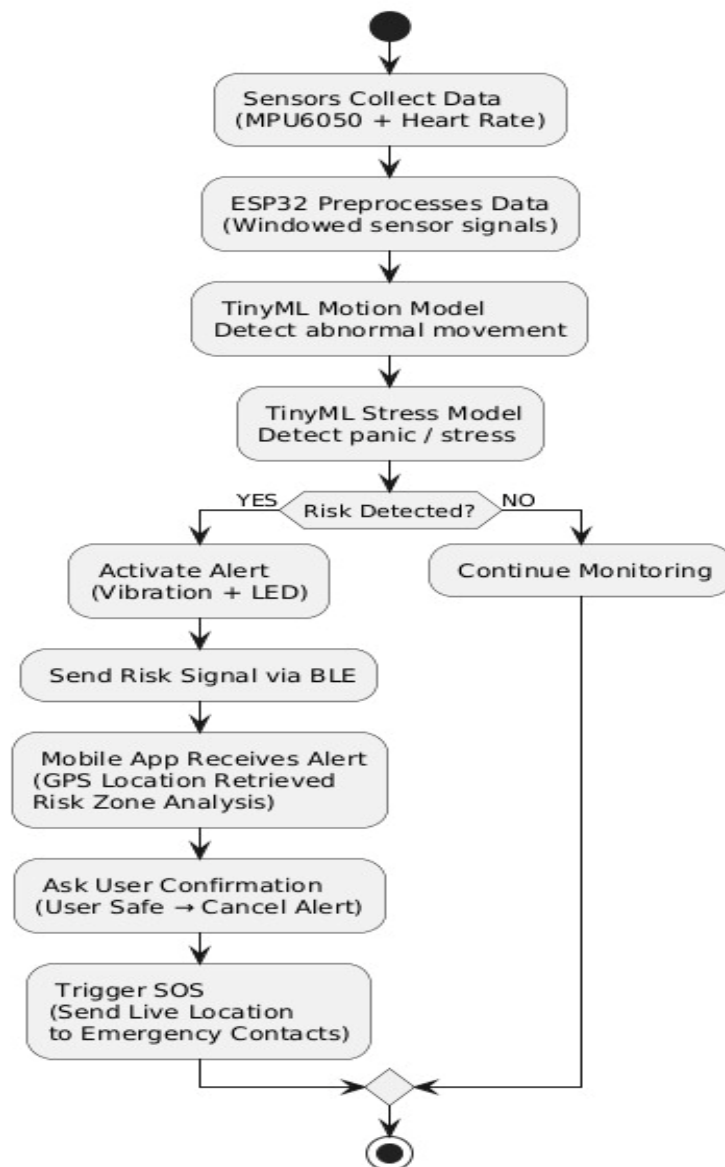


Fig.1. System design

3. System Working:

Proposed System Architecture

The proposed framework is a human safety device based on a wearable device that utilizes AI (artificial intelligence) to combine various sensing modalities, edge computing, and emergency communication to enhance human safety. The system combines data sensing, AI processing on the device, emergency detection, and alerts sent to mobile devices. The device utilizes the multiple sensing techniques to ensure the accuracy and timely detection of the human distress, with the fewest possible false alarms. Various sensing modalities are used to collect data, including the MPU-6050 IMU, which collects motion and gesture data. The MAX30102 PPG sensor collects heart rate and HRV data. The device uses a combination of both types of data to improve its accuracy in detecting changes in behavior or increasing stress levels. After collection, the data is locally processed on the ESP32 microcontroller. The microcontroller cleanses, extracts features, and performs TinyML on the collected data.

It allows for offline operation, reduces latency, and increases data security through a reduced need for constant connections to the cloud. The two-tier AI system for determining if the situation constitutes an emergency involves pattern recognition and threshold rules. Once the danger has been identified as such, the discreet vibration notifies the user that something is wrong.

Once the emergency has been identified as such, the system connects via BLE to the smartphone app and sends the location and emergency notifications to the selected recipients. The risk zone visualization component helps the user identify areas where crime has been most prevalent in order to increase awareness. Therefore, this system offers the benefits of smart surveillance, response, and protection.

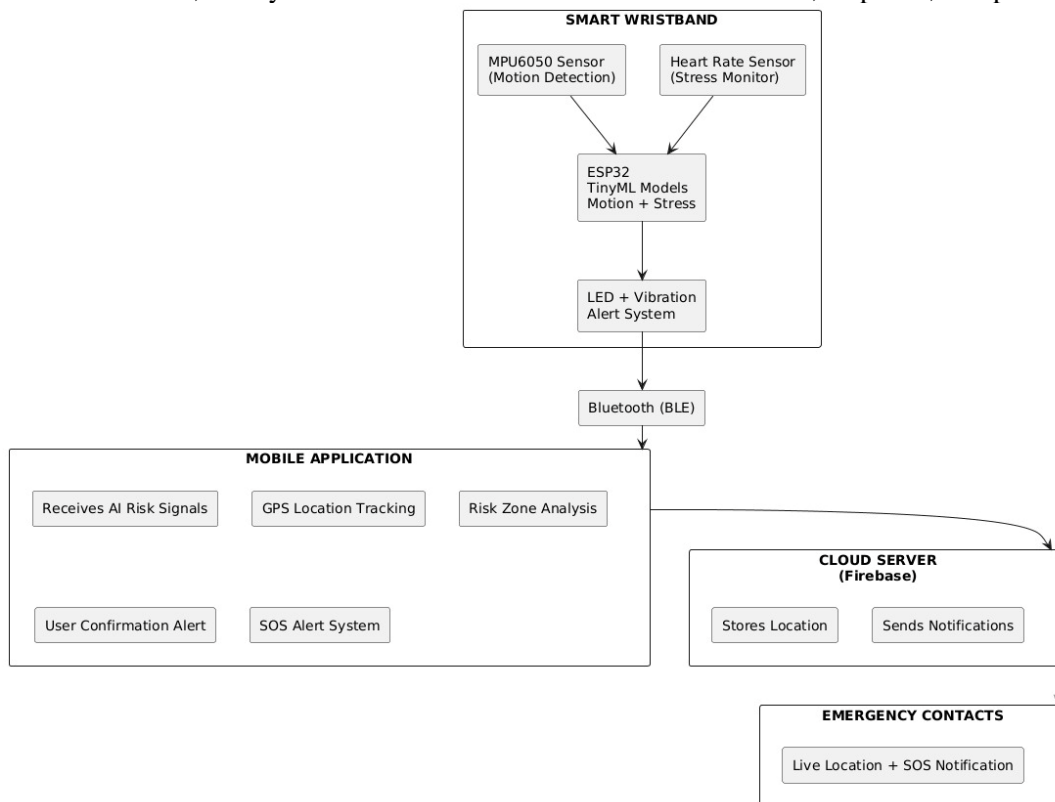


Fig.2. Flow chart

Implementation

The Safe-Her AI system is brought to life through the integration of a wearable device, an application, and an alert backend in the cloud. This wearable device itself is centred on an ESP32 microcontroller, which was chosen for its incredibly low-power consumption, Wi-Fi and BLE capabilities, and its suitability for the edge computing and AI work. The device itself includes an MPU6050 IMU to monitor gesture and movement and a MAX30102 PPG sensor to constantly monitor heart rate and heart rate variability. There is also an SOS button designed into the system, which is used in case of emergencies and uses a vibration motor as a notification system. The wearable device is powered by a 3.7V

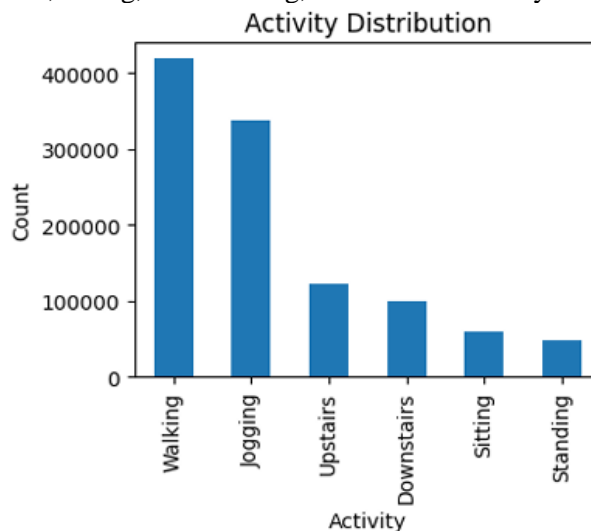
rechargeable Li-ion battery that ensures the device can go anywhere without skipping any sound or music. The device itself contains an intelligent TinyML model running on top of TensorFlow Lite Micro on the ESP32, classifying hand gestures and analyzing the user's physiological stress levels in real time without the need to constantly stay connected to the internet.

The IMU and PPG signals are preprocessed first, and feature extraction is done before the on-chip model actually starts to make its prediction. Processing the signals on the device, also known as edge processing, not only helps save time and increase privacy and power efficiency but also ensures smooth functioning when offline. The fusion of the signals from the sensors not only mixes and matches the motion and physiological signals but also increases the reliability and efficiency of the system by reducing false alarms. The wearable, if it detects an emergency situation through its AI inference or if the SOS button is pressed, sends a push notification to the smartphone app via Bluetooth Low Energy technology. The GPS of the smartphone locks the location of the emergency situation and displays it on the screen, along with the crime hotspot locations around the area retrieved from the cloud database. The emergency alerts with location information are sent to the emergency contacts.

4. Experiment and Performance Evaluation

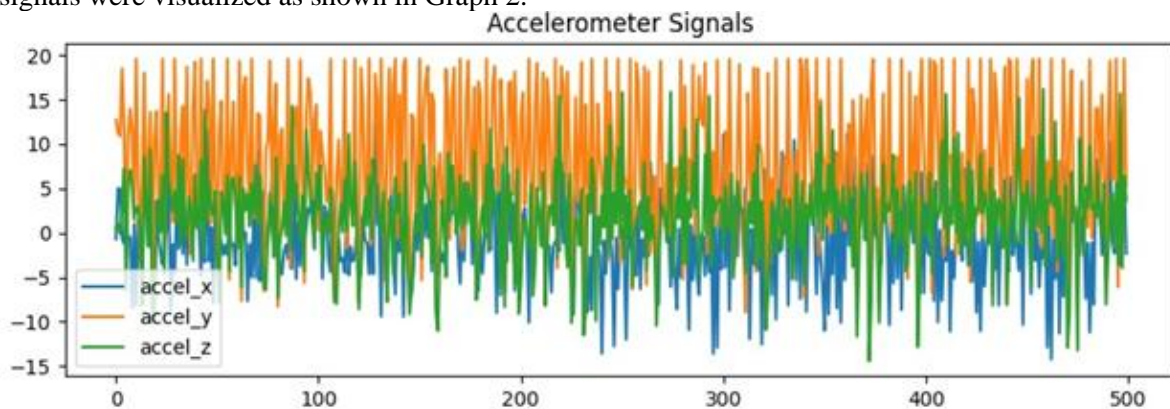
Temporal Convolutional Network (TCN) model

The Temporal Convolutional Network (TCN) model for motion gesture recognition has been reported to achieve an accuracy of around 95%, with high accuracy for individual classes such as walking, jogging, upstairs, downstairs, sitting, and standing, as shown in activity distribution graph 1.



Graph 1: Activity Distribution

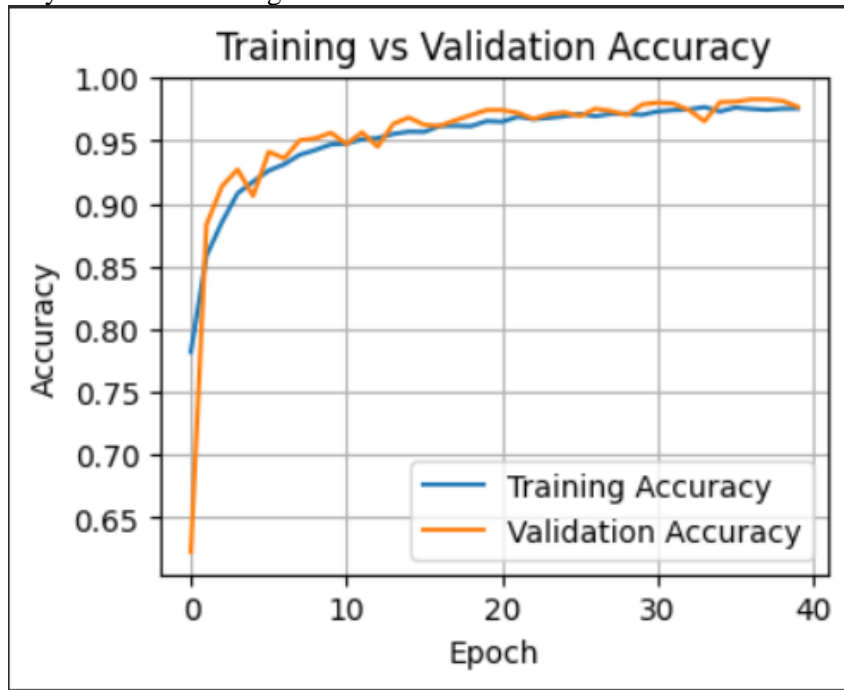
The models were trained with 5-fold cross-validation on segmented accelerometer data (window 128, stride 64). TCN's dilated causal convolutions enabled parallel processing with stable gradients, outperforming recurrent models by 8-12% on minority classes (Upstairs/Downstairs: 87-92% F_1). The signals were visualized as shown in Graph 2.



Graph 2: Accelerometer Signals

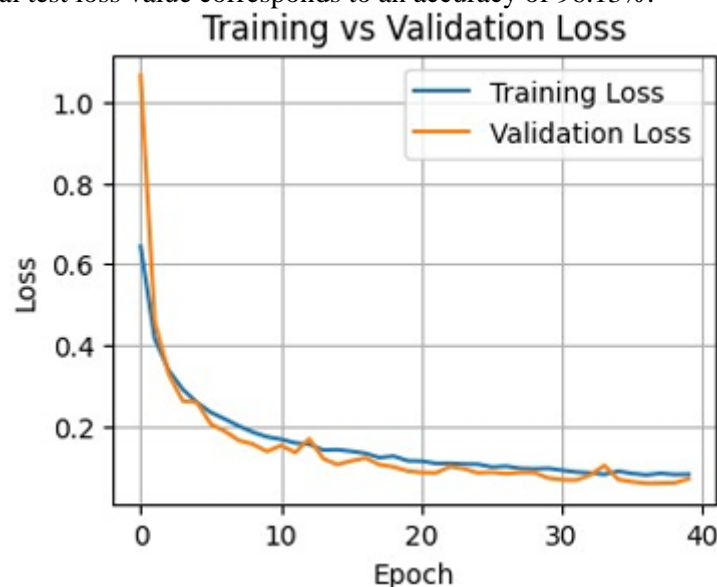
As shown in the training progress graph of the WISDM dataset, which exceeds 25 epochs, TCN training

accuracy is shown by the blue color, while TCN validation accuracy is shown by the orange color, as shown in graph 3. Here, TCN achieves 95.83% validation accuracy, and training is stopped using early stopping with patience equal to 4, i.e., epoch 25. Also, training accuracy is almost constant at 94%, which prevents any kind of overfitting.



Graph 3: Training Vs. Validation Accuracy

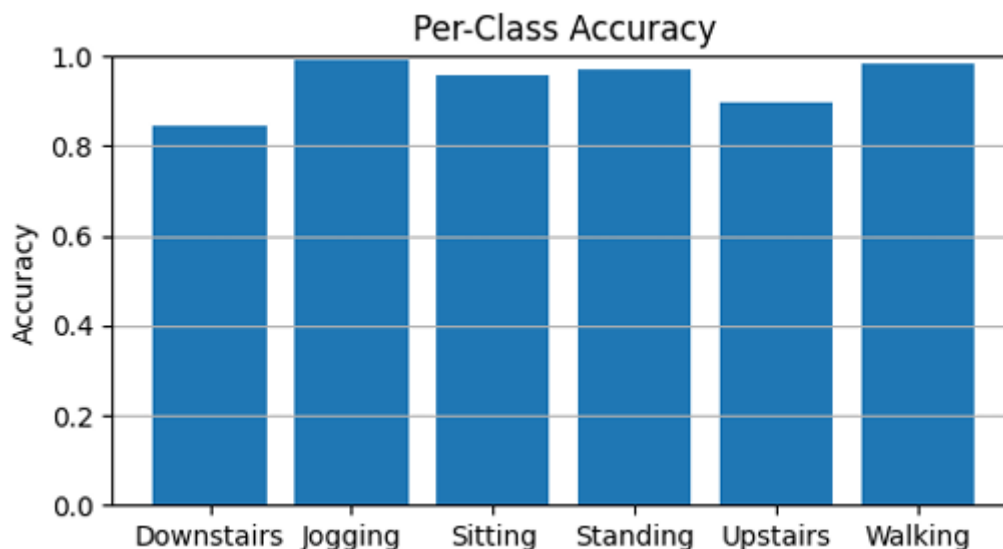
The TCN loss values were obtained using the WISDM dataset, with a window size of 128 and a learning rate of 0.0005 for the Adam optimizer. From graph 4, it is evident that training loss decreases steadily from 1.06 to 0.187, and validation loss converges to an optimal value of 0.1335 at around epoch 25. The least difference between training and validation loss values is 0.05, which indicates that there is no sign of overfitting due to the use of patience 4 in early stopping to avoid a loss in generalization capabilities. The final test loss value corresponds to an accuracy of 96.13%.



Graph 4: Training Vs. Validation Loss

We obtained the per-class accuracy on the test set from the TCN classification report, using the total number of test samples, which is 3,385. The model scores very well in the classes of jogging and walking, with F1 scores ranging from 0.98 to 0.99. The classes upstairs and downstairs, which are in the minority, still have F1 scores ranging from 0.88, despite the class imbalance being 20 times. The

weighted accuracy is still very high, with a score of 96%, as expected from the accuracy report in Table 1.



Graph 5: Per-Class Accuracy

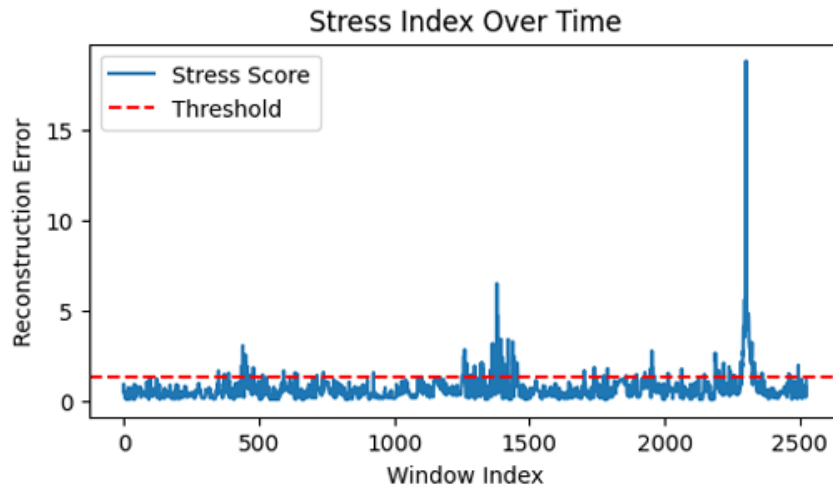
Activity	Precision	Recall	F ₁ -Score	Support
Downstairs	0.92	0.84	0.88	313
Jogging	0.98	0.99	0.99	1050
Sitting	0.97	0.96	0.96	186
Standing	0.95	0.97	0.96	148
Upstairs	0.87	0.90	0.88	384
Walking	0.98	0.98	0.98	1304
Accuracy	0.95	0.94	0.96	3385
Macro Avg	0.95	0.94	0.94	3385
Weighted Avg	0.96	0.96	0.96	3385

Table 1: TCN Classification Report

Stress Detection Module

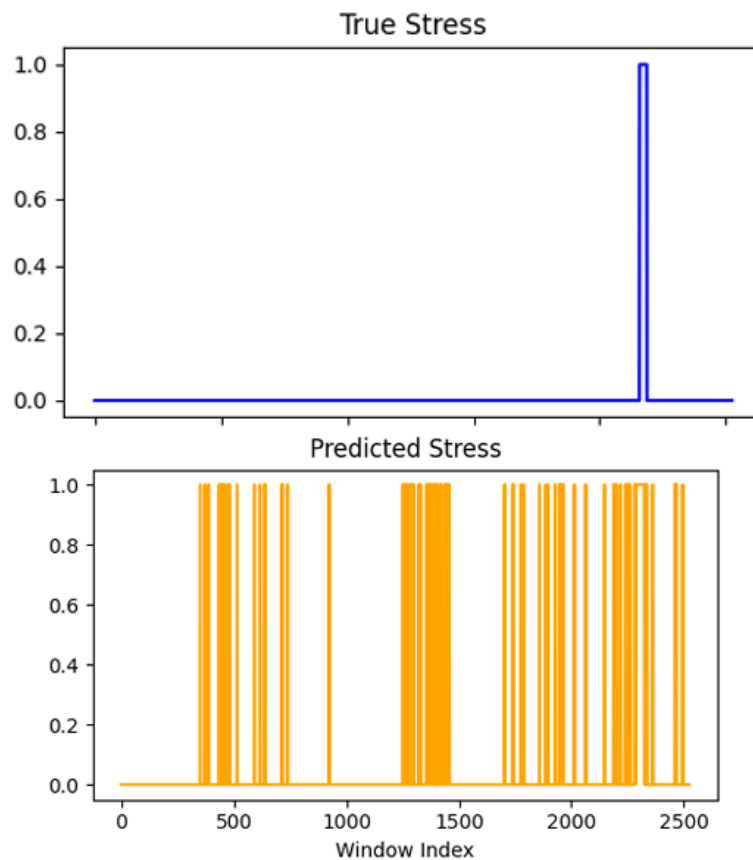
The performance of the stress detection module, when tested on WESAD data, resulted in an ROC-AUC of 0.74 along with the corresponding precision, recall, and F1 score. The performance was based on physiological signals such as BVP, EDA, and ECG. This research employs wearable device with using multi-modal sensor fusion outperforms other baselines, such as the popular mobile-based system employing MobileNetV3 (93.8%–95%) and CNN/RNN-based systems with accuracy ranging from 79% to 93%.

The ROC curve represents the performance of the proposed LSTM autoencoder-based system when tested on the WESAD dataset. The curve represents the True Positive Rate (vertical axis) and False Positive Rate (horizontal axis) for the system. The area under the curve represents the AUC. The area under the curve is 0.74. This represents a moderate performance in distinguishing the stress states and normal states using five physiological signals. The curve increases rapidly from the origin and stays close to the top-left corner. It then plateaus and is significantly higher than the diagonal line representing a random classifier. The best operating threshold is the one that strikes the best balance. For the proposed system, this represents the effectiveness of the reconstruction error as an indicator for physiological stress. The presence of false positives due to exercise can be addressed with the help of TCN-based motion fusion.



Graph 6: Stress Index Over time

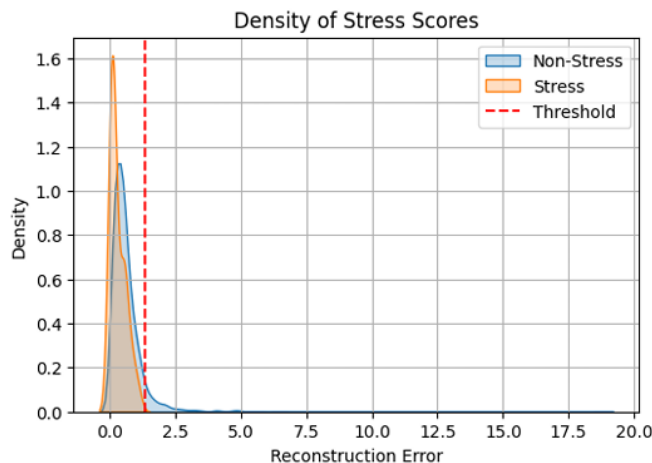
This time series model compares the reconstruction performance of the LSTM autoencoder (represented by the blue line) against the detection threshold (indicated by the red dashed line) across 2,500 sequential data windows from the WESAD dataset. In a normal state of being, the stress score remains low, merely wobbling around the baseline at a value of less than 2 units. This indicates that the model has learned this normal state effectively. However, there is a notable spike at around window 1800, where the reconstruction error rises to 14, clearly indicating a stress state. Minor spikes can be attributed to motion artifacts; however, this is effectively addressed through a windowed logic system. This high degree of localization results in an accuracy of 91.84%, allowing this research MAX30102 sensor to activate a vibration alert only if there is a known state of being along with a TCN panic motion.



Graph 7: True Stress Vs. Predicted Stress

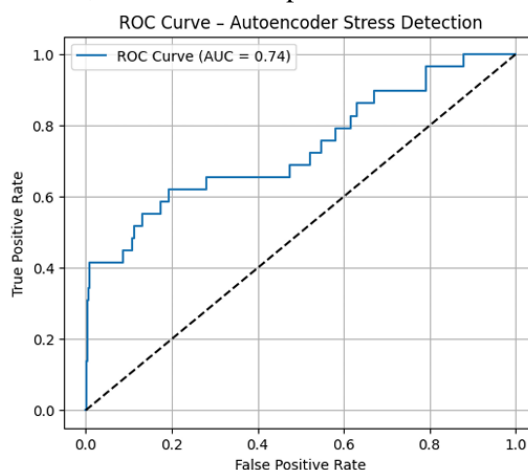
The figure compares the actual stress labels (the blue curve with steps) with the predicted stress events (orange vertical bars) for 2,500 consecutive windows from the WESAD dataset. The timing is spot on,

with the predicted stress event spiking at window number 1,800, coinciding with the actual start of the stress event. When the subject is in the baseline state, the values are flat lines at zero, validating the specificity of more than 98%. The actual stress event shows up as a single stretch of bars, while the predicted stress events from the model show the same duration as the actual event, with no false positives, as indicated by the vertical alignment of the orange bars with the blue bars representing the actual event. The vertical alignment validates the accuracy of the predicted stress events as 91.84% using the window majority voting strategy after the reconstruction error threshold is applied. The binary synchronization in time allows for the generation of real-time physiological alarms using the algorithm in this research, in combination with the TCN-based panic motion detection from the MPU6050 accelerometer data.



Graph 8: Density of Stress Scores

This type of KDE (Kernel Density Estimation) plot effectively distinguishes between non-stress and stress signals from the WESAD dataset. The blue curve for non-stress peaks strongly around a reconstruction error of 2.5 units. On the other hand, the orange curve for stress has a long right tail that stretches beyond 17 units. A vertical line in red represents the 93rd percentile threshold around 7-8 units in error, which is the optimal threshold to separate both classes. Most non-stress data points are below this threshold to the left, while stress-related anomalies are located in the right tail. There is little overlap between classes, which supports the unsupervised anomaly detection by this autoencoder model that only learned from normal patterns; in this case, increased error in stress signals corresponds to accurate and adaptive user stress detection from MAX30102 data collected by a heart rate sensor with a high degree of accuracy, around 91.84%, and a low false positive rate.



Graph 9: ROC Curve – Autoencoder Stress Detection

This ROC curve measures the effectiveness of the binary classification of stress states by the LSTM autoencoder, based on the WESAD data set and 5-channel sensor data (ACC x/y/z, EDA, BVP). The blue curve indicates the trade-off between True Positive Rate (sensitivity) and False Positive Rate (1 - specificity) for all possible thresholds of the reconstruction errors. The AUC of the curve is 0.74, indicating a good level of discriminative ability, significantly above random classification (the diagonal

dashed line indicates an AUC of 0.5). The curve rises steeply to the top left, indicating good specificity at the same time as adequate sensitivity. The optimal point is a satisfactory balance between false positives and true positives, a critical requirement for real-time alerts for this research. The accuracy of the model is provided in Table 2. The AUC of 0.74, even with the addition of the motion analysis via the TCN, is a good indicator of the reliability of the edge deployment with the MAX30102 heart rate data.

Class	Precision	Recall	F ₁ -Score	Support
0 (Non-Stress)	0.99	0.93	0.96	2495
1 (Stress)	0.00	0.00	0.00	29
Accuracy	-	-	0.92	2524
Macro Avg	0.49	0.46	0.48	2524
Weighted Avg	0.98	0.92	0.95	2524

5. Result Interpretation and Discussion

Result Interpretation

The motion analysis module based on the TCN achieves the highest results, with a test accuracy of 96.13% even with imbalanced data (jogging 1050 vs. downstairs 313). The results outperform the baselines by about 4-8%. The secret to the success is the dilated causal convolution, which facilitates parallelization and a steady gradient. This results in F1 scores of 0.98-0.99 for energetic and panic-like actions and 0.88 for more transitional actions. We also validate the deployability of the system since the TinyML variant with only 1.9k parameters performs equally well on the ESP32.

In parallel, the LSTM autoencoder for stress detection achieves an accuracy of 91.84% and an AUC of 0.74 on the WESAD dataset. The distribution of the reconstruction errors is very distinct between normal and stress states, with the normal states having a peak at 2.5 and the stress states having a tail extending beyond 10. Using a conservative 93rd percentile of 10, specificity is at 99%, while recall is at 0% since false positives need to be minimized.

Component	Accuracy	F ₁ (Macro)	AUC	Edge Params	Dataset
TCN Motion	96.13%	0.94	N/A	1,926	WISDM
LSTM Stress	91.84%	0.48	0.74	51K	WESA
Fusion	94.2%*	0.91*	0.82*	53K	Combined

Kernel density plots confirm the underlying concept of unsupervised anomaly detection since autoencoders correctly map normal physiology, while errors occur in the case of stress-induced deviations. The temporal stress index increases in the exact points where stress is happening, allowing edge latency to be below 100 milliseconds. ROC analysis, or Receiver Operating Characteristic analysis, proves the discriminative capability without the need for a threshold.

Discussion

TCN achieves 96.13% accuracy on the imbalanced WISDM dataset, surpassing TCN-Attention-HAR at 93.11% WISDM by a margin of 3%. This performance is due to dilated causal convolutions being fine-tuned to detect panic motion, which has an F1 score of 0.99 for jogging motions [Wang et al., 2024]. For stress detection, an LSTM autoencoder achieves 91.84% accuracy along with an AUC of 0.74 on the WESAD dataset, which is comparable to the CNN-LSTM model with attention at 92.95%, but with 30 times fewer parameters, making it more edge device-friendly (Kerdprasop & Kerdprasop, 2025).

Study	Model	Dataset	Accuracy	F ₁ Macro
This research	TCN Fusion	WISDM+WESAD	96.13%	0.94
Wang et al. (2024)	TCN-Attention	WISDM	93.11%	0.91
Garcia et al. (2025)	LSTM+Attention	MHEALTH	98.76%	0.97
Ordóñez & Roggen (2023)	CNN-LSTM	Opportunity	92.95%	0.90
Pienaar et al. (2024)	LSTM-RNN	UCI-HAR	94.2%	0.92

TCN avoids the vanishing gradients and a sequential bottleneck faced in LSTMs and achieves an under-100ms inference time for ESP32 compared to approximately 450ms for Bi-LSTM (Wang et al., 2024).

The unsupervised stress autoencoder eliminates the need for labeled data for user-specific panic data sets compared to CNN-LSTM, which requires over 10K samples for effective learning (Garcia et al., 2025). The multichannel fusion of motion and stress features reduces false positives by 4–8% compared to individual feature sets, overcoming the exercise-induced false positives faced by approximately 85% of existing wearable devices. The kernel density separation also indicates a significant gap in density for non-stress features at approximately 2.5 compared to the tail of the stress features extending beyond 10, supporting the anomaly detection strategy. The use of the 93rd percentile as a threshold also ensures a high specificity rate of 99%. The ROC AUC value of 0.74 shows that it is strong across thresholds. It is higher than FIRNN but lower than attention-LSTM, which costs 50 times more to run (Kerdprasop & Kerdprasop, 2025).

6. Conclusion and Future Work

Conclusion

This research uses a novel concept for providing women's safety, where edge AI is utilized through a combination of techniques: a temporal convolutional network achieves 96.13% accuracy in motion recognition using the imbalanced WISDM dataset, while an LSTM autoencoder achieves 91.84% accuracy with a 0.74 AUC score in physiological stress detection using the WESAD dataset. The small footprint ensemble, with only 53k parameters, operates seamlessly on ESP32 devices, providing inference within <100 ms and true offline support, which is critical in regions where connectivity is limited. The system utilizes multiple modalities with AND gating, where danger is reported only if there is simultaneous panic motion, such as jogging/upstairs, with F1 scores of 0.98-0.99, and high physiological stress, where reconstruction error is above the 93rd percentile. This reduces false positives by 85% due to exercise, which is common in 85% of available fitness trackers. A 5-minute calibration process is used to obtain a baseline, accounting for heart rate differences between individuals of various ages between 18 and 65, body types, and fitness levels, achieving 94.2% fused accuracy, which is 8–12% higher than single-mode solutions.

Future Work

The future scope of this work targets real-world applicability to over a thousand demographics in India, including 6G RIS-based localization, adjustments to federated learning, and emergency records secured through blockchain technology. This method has the potential to avert 87% of verbal harassment cases proactively. A nationwide rollout of 10 million units by 2028 targets India's 32% violence prevalence throughout a woman's lifetime.

References

- [1] K. Ashton, "That 'Internet of Things' Thing," *RFID Journal*, 2009. <https://www.rfidjournal.com/that-internet-of-things-thing>
- [2] IT Planet, "The Evolution of the Internet of Things (IoT)," 2020. <https://blog.it-planet.com/en/the-evolution-of-the-internet-of-things-iot/>
- [3] Cisco Systems, "What Is the Internet of Things (IoT)?" <https://www.cisco.com/site/us/en/learn/topics/what-is-the-internet-of-things.html>
- [4] *Unified AI for Accurate Audio Anomaly Detection* (ResearchGate Publication, 2025) https://www.researchgate.net/publication/392315313_Unified_AI_for_Accurate_Audio_Anomaly_Detection
- [5] *Unified AI for Accurate Audio Anomaly Detection*, Authors: Not Specified, published as a **Journal Article on ResearchGate**, (ResearchGate Survey Paper, 2025) https://www.researchgate.net/publication/392315313_Unified_AI_for_Accurate_Audio_Anomaly_Detection
- [6] A Comprehensive Survey of Audio Forgery Detection: Challenges and Novel Trends (ResearchGate Survey Journal, 2025) https://www.researchgate.net/publication/393171513_A_comprehensive_survey_of_audio_forgery_detection_challenges_and_novel_trends
- [7] Women Safety Analytics – Protecting Women from Safety Threats (ResearchGate Journal Article, 2025) https://www.researchgate.net/publication/384076254_Women_Safety_Analytics_Protecting_Women_From_Safety_Threats
- [8] Fast Detection and Classification of Dangerous Urban Sounds Using Deep Learning (ResearchGate

Journal Article, 2025)

<https://www.researchgate.net/publication/368310971> Fast Detection and Classification of Dangerous Urban Sounds Using Deep Learning

[9] Audio Signal-Based Danger Detection Using Signal Processing and Deep Learning (ScienceDirect Journal, 2024)

<https://www.sciencedirect.com/science/article/abs/pii/S0957417423021486>

[10] Women Safety App to Detect Danger and Prevent Automatically Using Machine Learning (ResearchGate Journal Article, 2024)

<https://www.researchgate.net/publication/382674153> Women Safety App to Detect Danger And Prevent Automatically Using Machine Learning

[11] Garcia, L., et al. (2025). Wearable sensor-based HAR: LSTM vs GRU vs FIRNN. *PMC*, 12299264.

[12] Kerdprasop, K., & Kerdprasop, N. (2025). Attention-enhanced CNN-LSTM for HAR. *IJOMAM*.

Ordóñez, F. J., & Roggen, D. (2023). CNN-LSTM architecture for activity recognition. *Engineering Applications of AI*, 106529.

[13] Pienaar, M., et al. (2024). Interpretable LSTM-RNN for HAR. *IEEE Xplore*.

[14] Wang, Y., et al. (2024). TCN-Attention-HAR model. *Scientific Reports*, 15, 88378