Grenze International Journal of Engineering and Technology, June Issue

Edge-AI Wearable for Real-Time Health Monitoring, Intelligent Medical Analysis, and Proactive Alert Management

Madan Kumar G[1], Mohith Gonthukokala[2], Mohamad Fahad Abdulla[3], Shanthu Kumar M[4], M Roopa[5] and Pavithra G[6]
[1-6]Dayananda Sagar College of Engineering, Bengaluru, India.
Email: dr.pavithrag.8984@gmail.com

***Abstract*— Access to continuous health monitoring remains limited due to high medical infrastructure costs and the restricted capabilities of conventional wearable devices. This work presents an integrated wearable health monitoring system combining on-device processing with cloud-based intelligence. The system is built on an ESP32-S3 platform and incorporates multiple sensors, including MAX30102 for heart rate and SpO_2, MLX90614 for non-contact temperature sensing, MPU6050 for motion tracking, and a galvanic skin response sensor for stress detection. A TCA9548A I^2C multiplexer ensures reliable communication, while an OLED display provides real-time feedback.The wearable performs local processing using threshold-based methods to detect anomalies in vital signs and activity patterns, enabling immediate alerts through vibration without requiring network connectivity, thereby ensuring faster response and improved data privacy. Simultaneously, the device connects to a Node.js-based backend using WebSocket communication for real-time data transfer and monitoring. Emergencyalerts are enhanced with location tracking and SMS notifications containing live map links.The system also supports AI-based medical data analysis, allowing users to upload reports such as X-rays or ECGs for structured interpretation and recommendations. A React-based interface provides visualization, reminders, and additional services, demonstrating a cost-effective and scalable solution for real-time health monitoring**

***Index Terms*— Wearable Health Monitoring, ESP32-S3, Edge AI, Large Language Models, Groq API, MAX30102, MLX90614, MPU6050, Galvanic Skin Response, Real-Time Anomaly Detection, Medical Report Analysis, Twilio SMS, Socket.io, React, TCA9548A.**

I. INTRODUCTION

The increasing prevalence of non-communicable diseases, including cardiovascular conditions, respiratory disorders, metabolic syndromes, and stress-related illnesses, continues to place significant pressure on healthcare systems worldwide. Limited access to medical infrastructure and uneven geographic distribution of healthcare services further complicate timely diagnosis and treatment. According to global health estimates, a large proportion of premature deaths associated with these conditions could be prevented through early detection and continuous monitoring. However, outside clinical environments, such monitoring remains difficult to achieve in a practical and affordable manner.

Grenze ID: 01.GIJET.12.2.817_26

Most commercially available wearable devices, such as smartwatches and fitness bands, provide basic measurements like heart rate and step count. While useful, these devices lack the ability to interpret medical data, analyze images, or provide meaningful health guidance. They do not support features such as medical report analysis, skin condition identification, or intelligent responses to user health queries. As a result, users often receive raw data without sufficient context, limiting its usefulness for informed decision-making.

Recent advancements in embedded systems and artificial intelligence have created new opportunities to address these limitations. Modern microcontrollers, including the ESP32-S3, offer sufficient processing capability and energy efficiency to support real-time data analysis directly on wearable devices. At the same time, large language models accessible through low-latency APIs enable rapid interpretation of both textual and visual medical data. In addition, freely available APIs for messaging, location services, and mapping have made it possible to build connected healthcare solutions without significant infrastructure cost.

This work presents an AI-enabled wearable health monitoring system that combines these developments into a unified platform. The system integrates a multi-sensor wearable device with on-device anomaly detection, a cloud-connected backend, and an interactive web application. It supports continuous physiological monitoring, AI-assisted medical analysis, conversational health interaction, reminder management, and emergency alerting with location support. Together, these components provide a practical and accessible solution for real-time health monitoring and preventive care

II. Scope and Objectives of the Work

In this section, the scope & objectives of the works are presented in a nutshell.

The scope of the proposed system encompasses the design and development of a cost-effective, AI-powered wearable health monitoring solution that integrates multi-sensor data acquisition, edge processing, and cloud-based services. It covers real-time monitoring of physiological parameters such as heart rate, SpO_2, body temperature, motion, and stress levels, along with anomaly detection and immediate alert generation. The system extends to cloud connectivity for data storage, visualization, and advanced analytics, including AI-based medical report interpretation. It also includes user-centric features such as reminder systems, emergency notifications with geolocation, and a web-based interface for health management. The solution is applicable across domains such as remote patient monitoring, Objectives

- To design and implement a wearable health monitoring system using ESP32-S3 with integrated multi-sensor support.
- To enable real-time physiological data acquisition and on-device processing using edge computing techniques.
- To develop efficient anomaly detection mechanisms for immediate alerts without dependency on network connectivity.
- To integrate cloud-based backend services for data storage, remote monitoring, and advanced analytics.
- To implement AI-based analysis for medical reports and health insights.
- To provide a user-friendly interface for visualization, reminders, and personalized health recommendations.
- To ensure low cost, scalability, energy efficiency, and enhanced data privacy in healthcare monitoring.

III. Proposed Methodology and Block Diagram

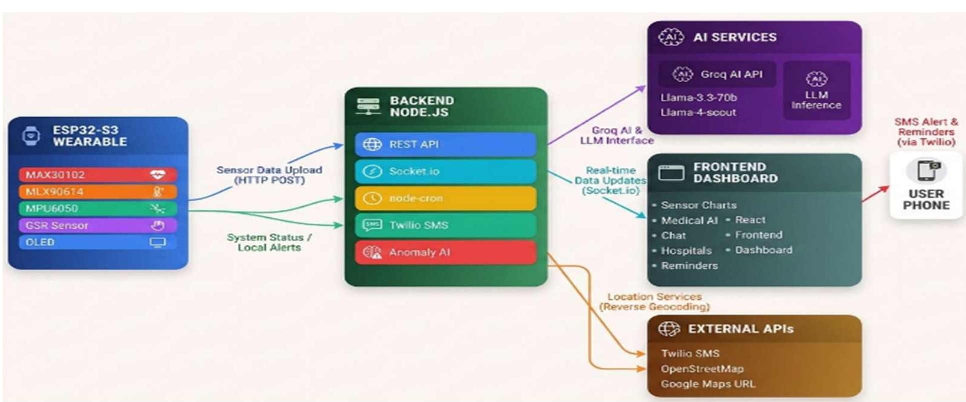

Fig. 1: Architectural flow diagram of the yolomart

A. Hardware Components

The wearable device is designed around the ESP32-S3 N8R2 module, featuring a dual-core Xtensa LX7 processor, 8 MB external OPI PSRAM, and 8 MB flash memory, along with integrated Wi-Fi and hardware-accelerated cryptographic support. Its compact architecture makes it well suited for wrist-worn health monitoring applications. To manage multiple sensors sharing the I²C bus and avoid address conflicts, a TCA9548A eight-channel I²C multiplexer (address 0x70) is employed, enabling channel-isolated communication by assigning each sensor to a dedicated channel.

Physiological sensing is achieved through multiple integrated modules. The MAX30102 pulse oximeter (channel 0, address 0x57) measures heart rate and SpO_2 using photoplethysmography with red (660 nm) and infrared (880 nm) LEDs. The MLX90614 infrared thermometer (channel 1, address 0x5A) enables non-contact body temperature measurement by detecting emitted infrared radiation from the skin within a range of 1–3 cm. Motion and activity tracking are handled by the MPU6050 inertial measurement unit (channel 2, address 0x68), providing three-axis acceleration (±2g) and angular velocity (±250°/s) for step detection and movement classification.

A 128×64 OLED display based on the SSD1306 controller (channel 3, address 0x3C) presents real-time sensor readings, alerts, and reminders. Additionally, a galvanic skin response sensor is connected to GPIO 4 to monitor skin conductance, indicating stress levels. A coin vibration motor, driven via GPIO 5 through an NPN transistor with a 1 kΩ base resistor and 1N4007 diode, provides haptic alerts.

All components are assembled on a breadboard prototype with structured interconnections, demonstrating a compact, modular, and scalable wearable health monitoring system.

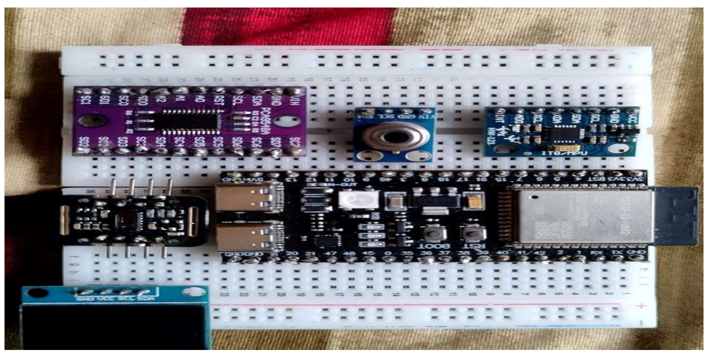

Fig. 2: Actual hardware prototype showing ESP32-S3 development board, TCA9548A I2C multiplexer (purple), MLX90614 IR temperature sensor (blue round), MPU6050 IMU (blue rectangular, top-right), additional sensor module (left), and SSD1306 OLED display (bottom-left) assembled on breadboard

TABLE I. COMPARISON BETWEEN CONVENTIONAL WEARABLES AND THE PROPOSED AI HEALTH MONITOR

Parameter	Conventional Wearable	Proposed AI Health Monitor
Sensors	Single (HR only)	MAX30102, MLX90614, MPU6050, GSR
AI Analysis	None	Groq LLM — medical reports, skin, chat
Anomaly Detection	Manual threshold only	Edge AI on ESP32-S3 + server-side ML
Emergency Alert	App notification only	SMS + live GPS location via Twilio
Reminders	Not supported	Water + medicine — OLED, vibration, SMS
Medical Reports	Not supported	AI analysis of X-rays, blood tests, ECG
Real-time Dashboard	Basic app only	Full web app — charts, maps, AI planner
Hospital Search	Not supported	OpenStreetMap — nearby hospitals & specialists

TABLE II: HARDWARE COMPONENTS AND INTERFACE SPECIFICATIONS

Component	Interface	Parameter Measured	Role
ESP32-S3 N8R2	–	–	MCU + Edge AI
TCA9548A Mux	I2C (0x70)	–	I2C bus isolation
MAX30102	I2C Ch0 (0x57)	Heart rate, SpO2	Pulse oximetry
MLX90614	I2C Ch1 (0x5A)	Body temperature	IR thermometry
MPU6050	I2C Ch2 (0x68)	Accel, Gyro (6-axis)	Motion & step detect
SSD1306 OLED	I2C Ch3 (0x3C)	–	128×64 display
GSR Sensor	Analog GPIO4	Skin conductance	Stress monitoring
Vibration Motor	Digital GPIO5	–	Haptic alerts

The block diagram illustrates a wearable health monitoring system built around the ESP32-S3 microcontroller, which acts as the main processing unit. The system integrates multiple sensors, including MAX30102 for heart rate and SpO_2, MLX90614 for temperature measurement, MPU6050 for motion tracking, and a GSR sensor for stress analysis. A TCA9548A I^2C multiplexer is used to manage communication between these sensors efficiently. The device continuously collects and processes data in real time, applying lightweight on-device models to identify abnormal conditions. The results are displayed on an OLED screen, and alerts are provided through a vibration motor when needed. Wireless connectivity allows the system to share data with external platforms for monitoring, while still maintaining fast response and user data privacy.

B. Operational Logic

The ESP32-S3 firmware executes a continuous sensing loop at approximately 25 Hz. On each cycle, the TCA9548A channel selector is activated sequentially to read the MAX30102, MLX90614, and MPU6050 sensors; the GSR sensor is sampled directly via the ADC; and the SpO2 and heart rate values are extracted from the MAX30102 IR and red photodiode buffers using the MAXIM SpO2 algorithm. The accelerometer magnitude vector is computed and compared against a threshold of 1.2g with a 250 millisecond debounce window for step counting.

All sensor values are then evaluated against the anomaly thresholds in Table III. If any threshold is exceeded, the firmware drives the vibration motor with three short pulses and updates the OLED display with an alert screen. Sensor data is serialized to JSON and posted to the backend every two seconds, and the pending reminder queue endpoint is queried every five seconds.

When a reminder item is present in the queue, the ESP32 displays the reminder message on the OLED for eight seconds and produces a type-specific haptic pattern —two gentle pulses for water reminders and three longer pulses for medicine reminders.

TABLE III: EDGE AI ANOMALY DETECTION THRESHOLDS

Vital Parameter	Sensor	Normal Range	Alert Threshold
Heart Rate	MAX30102	60 – 100 BPM	< 50 or > 120 BPM
SpO2	MAX30102	≥ 95%	< 95%
Body Temperature	MLX90614	36.1°C – 37.2°C	< 35°C or > 38.5°C
GSR (Stress)	Analog ADC	0 – 700 raw units	> 800 raw units
Step Magnitude	MPU6050	< 1.2 g	Crossing 1.2 g = step

A. Backend Microservices

The backend is implemented using Node.js, Express, and Socket.io for real-time communication. It provides RESTful APIs for sensor data handling, reminders, SMS alerts, medical data processing, AI chat, and hospital search.

Sensor data from the ESP32 is normalized, analyzed for anomalies, and broadcast to clients. Reminders are scheduled using node-cron, while SMS alerts (via Twilio) include optional location-based Google Maps links. Medical uploads are processed using AI models for text and image analysis, and hospital data is retrieved using the OpenStreetMap Overpass API.

B. AI Inference Pipeline

AI processing is powered by the Groq API for both text and image analysis. Medical reports are summarized with identification of abnormalities and recommendations, while images are assessed for visible conditions and severity. A conversational AI assistant supports multiple health-related queries, providing continuous and contextual user interaction.

C. Frontend Application

The frontend, built with React, Vite, and Tailwind CSS, offers a real-time dashboard for physiological data visualization. It includes modules for medical report upload, skin analysis, nearby hospital search, AI chat, activity tracking, and personalized health planning. A reminder system supports daily health routines.

D. Summary of Results

The system demonstrates high accuracy, with heart rate (±3 BPM), SpO_2 (±1%), and temperature (±0.3°C). Step detection error is below 5%, and GSR effectively reflects stress variations. AI responses are generated within seconds, and end-to-end alert delivery occurs in under five seconds, ensuring reliable real-time performance.

IV. SOFTWARE ARCHITECTURE

TABLE IV: SYSTEM PERFORMANCE COMPARISON: PROPOSED SYSTEM VS. NON-AI APPROACH

Feature / Metric	Without AI Integration	Proposed System
Medical report interpretation	Manual doctor review (days)	AI summary in < 10 seconds
Skin condition assessment	Dermatologist appointment required	Image-based AI analysis instantly
Emergency alert with location	Voice call or manual message	Automated SMS with Google Maps URL
Diet plan generation	Nutritionist consultation (cost, time)	Personalized 7-day plan in < 30 s
Exercise plan	Trainer or generic online plan	4-week progressive AI plan, goal-aware
Hospital navigation	Search engine / manual inquiry	Real-time map with specialists shown
Sensor anomaly to alert latency	Not automated	< 3 seconds (edge AI + socket)
Reminder delivery channels	None / basic phone app	OLED + haptics + browser + SMS

B. AI Model and API Utilization

Table V presents the AI models used across different system features along with their average response times and cost considerations. The system is designed to operate within free-tier limits, making it practical for low-cost deployment.

The Groq API supports multiple requests per minute and a large number of daily requests without cost, while OpenStreetMap services are freely available for location-based queries. Twilio also provides initial credit that is sufficient for a large number of alert messages.

These factors show that the system can deliver meaningful AI-based health monitoring without requiring continuous infrastructure expenses.

TABLE V: AI MODELS, LATENCY AND COST SUMMARY

AI Feature	Model / API Used	Avg. Latency	Cost
Medical report (PDF)	Groq llama-3.3-70b-versatile	4 – 7 s	Free tier
Medical scan (image)	Groq llama-4-scout-17b (vision)	5 – 9 s	Free tier
Skin analysis	Groq llama-4-scout-17b (vision)	4 – 8 s	Free tier
Health chatbot	Groq llama-3.3-70b-versatile	1 – 3 s	Free tier
Diet / exercise plan	Groq llama-3.3-70b-versatile	6 – 12 s	Free tier
Hospital search	OpenStreetMap Overpass API	2 – 5 s	Free (no key)
SMS alert w/ location	Twilio REST API	< 5 s	Free trial

C. Reminder Delivery Architecture

The reminder system uses a backend-based scheduling approach with node-cron, eliminating the need for time management on the ESP32 and ensuring persistence across device restarts. The wearable periodically retrieves reminders, displays them on the OLED screen, and triggers vibration alerts. Simultaneously, notifications can be sent to the web interface and via SMS to caregivers. This multi-channel delivery (device, web, and SMS) improves reliability and ensures critical reminders are not missed.

D. Location-Enriched Emergency Alerting

The system enhances emergency alerts by integrating real-time location data. The frontend collects GPS coordinates using the Geolocation API and sends them to the backend via Socket.io. Upon detecting an anomaly, the backend appends a Google Maps link with the latest coordinates to the SMS alert sent via Twilio. This enables recipients to instantly access the user’s location. A rate limiter prevents excessive alerts during continuous abnormal conditions.

VI. CONCLUSIONS

In conclusion, the proposed AI-enabled wearable health monitoring system integrates multi-sensor data acquisition, edge-based anomaly detection, cloud analytics, and location-aware emergency alerting into a unified and cost-effective platform. The ESP32-S3 device performs real-time monitoring and on-device processing, while the backend ensures rapid communication and AI-driven health insights with alert latency under five seconds. The system supports medical analysis, reminders, and interactive assistance, making it suitable for personal and community use. Future work will focus on enhancing on-device intelligence using TinyML, improving connectivity, optimizing power consumption, and enabling privacy-preserving data-driven advancements.

ACKNOWLEDGEMENT

The authors would like to thank the following 6th Semester students of Dayananda Sagar College of Engineering for their contribution. Sachin,Sumeet,Naveen, Jatin,Krishna goyal,Manish kumar sahu, Tyrus

REFERENCES

[1] G. Saravanan, R. Punitha Gowri, Arvinfh S. V, Dheena R. S, and Krishna Kumar M, “Survey on Remote Monitoring System,” 2024 9th International Conference on Communication and Electronics Systems (ICCES).

[2] Li Tung Tan, Raihanah Syamimi Mohammad Shahrizal, Muhammad Naqib’Aqil Mohd Nadzri, Norashikin Yahya, and Suhaila Badarol Hisham, “Remote Patient Monitoring System,” 2022 IEEE 5th International Symposium in Robotics and Manufacturing Automation (ROMA).

[3] Md Ameenul Hasan and Megha P. Arakeri, “Remote Patient Monitoring System Using IoT and Artificial Intelligence: A Review,” 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC).

[4] K., Jyotish Chandra; Jishnu; Ashirwaad; Sarbesh; G.,” Flower Classification using Pre-trained ResNet Models in Computer Vision” Grenze International Journal of Engineering & Technology (GIJET), 2024, Vol 10, Issue 2,Part 6, p6737. https://www.scopus.com/record/display.uri?eid=2-s2.0-85209065920&origin=resultslist

[5] Shweta Goyal, Sachine Sharma, Pushpendra Singh Kharayat, Sugam Gupta-"E-Waste: Promoting the Need for Green Electronics", 2023 International Conference on Power Energy, Environment & Intelligent Control (PEEIC), IEEE, 2023.

[6] Varshith Raj C., "Hardware-Efficient Simulation and RTL-Based Implementation of SPECK Algorithm Using Verilog: A Solution for Secure and Fast VLSI Cryptography," International Conference on Emerging Trends and Innovations in ICT (ICEI), Pune, Maharashtra, India, 9-10 Jan 2026, pp. 1-7

[7] Saba Fatima, Kiruba M., P. Nisha "Enhancing Digital Image Processing with Energy-Efficient Posit Multipliers for VLSI based Quantum Computing Applications," The 7th IEEE International Conference on Information Systems and Computer Networks (ISCON-25), Department of Computer Engineering and Applications, GLA University, Mathura, Uttar Pradesh, India, Paper Id 51, 5-6 September 2025,

[8] Balaji G.H., "Advanced VLSI-Based Design of Multiple Path OTAs with Dual Flipped Voltage Follower for Signal Processing & Communication Applications," The IEEE International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), BNM Institute of Technology, Bangalore, Karnataka, India, 16-17 January 2025, pp. 1-5, doi: 10.1109/IITCEE64140.2025.10915364

[9] Rajesh Maharudra Patil, , “Harnessing QCA for High-Performance, Low-Power Arithmetic Circuits: A Focus on Multipliers and Square Circuits,” The Fourth International Conference on Mobile Networks and Wireless Communications (ICMNWC), SSIT, Tumakuru, Karnataka, India, Paper ID 506, pp. 1-8, 4-5 December 2024, Electronic ISBN:979-8-3315-2834-8

[10] N. Lakshmi, “CMOS Implementation of Multipath Fully Differential OTA with Dual Flipped Voltage Follower in 50 nm and 75 nm CMOS Technologies using Cadence Tool,” 2nd International Conference on Distributed Computing and Optimization Techniques IEEE ICDCOT-2024, SJB Institute of Tech., Bangalore, Karnataka, India, 15-16 March 2024, pp. 1-8.