ZANCO Journal of Pure and Applied Sciences

TYPE Original article
DOI http://dx.doi.org/10.21271/ZJPAS.38.3.16

OPEN ACCESS

*Corresponding author
Soran Bakhtyar Noori
soran.noori87@gmail.com

RECEIVED :30/07/2025
ACCEPTED :07/12/2025
PUBLISHED :30/06/ 2026

KEYWORDS:
Emotion recognition, IoT sensors, machine learning, mood measurement, wearable technology, mental health monitoring.

Cost-Effective Mood Detection Through IoT Wearable Sensors and Machine Learning Classification

Soran Bakhtyar Noori * & Dler Salih Hassan

Department of Computer Science and Information Technology, College of Science, Salahaddin University-Erbil, Erbil, Kurdistan Region, Iraq.

ABSTRACT

A cost-effective wearable Internet of Things (IoT) system is presented for real-time mood and stress detection using MAX30100 photoplethysmography, galvanic skin response (GSR), and MLX90614 wrist temperature sensors integrated with a Raspberry Pi edge device. Physiological signals are acquired, preprocessed on-device, segmented into 60 s windows with 50% overlap, and transformed into heart-rate variability, GSR, and temperature features. Six machine-learning classifiers are trained and evaluated with support from the WESAD dataset and a custom dataset collected from 25 participants, using a single stratified 70/30 train–test split and leave-one-subject-out cross-validation. Results indicate that low-cost hardware and lightweight models can achieve promising performance for real-time mood and stress monitoring, suggesting feasibility for mobile and ambulatory healthcare applications. Future work will extend deployment to a mobile companion app with real-time alerts, increase cohort size for stronger subject-independent validation, and explore compact deep models while preserving on-device privacy.

1.Introduction

Mental health disorders such as stress, anxiety, and depression are on the increase across the globe, with millions of cases worldwide, and the need to develop new methods of monitoring and control. Although traditional mental health assessment methods have some strengths, they tend to be limited by factors such as subjectivity, expense, lengthy processes, and inaccessibility, especially in remote or underserved communities. Recent technological developments in wearable technology, Internet of Things (IoT) infrastructure, and machine learning algorithms have opened up never-before-seen possibilities for continuous, objective, and affordable mental health monitoring (Pinge et al., 2024, Bhukya et al., 2023). There has been a phenomenal growth in the emotion detection and recognition market, which is estimated to reach USD 56.0 billion by 2024, representing a significant improvement from USD 21.6 billion in 2019 (Pal et al., 2021). Recent studies have shown that physiological signals can be used as reliable measures of emotional states, and that wearable sensors can also be used for continuous monitoring of these biomarkers in real-world settings (Pinge et al., 2024). One of the primary benefits of IoT integration in mental health monitoring is the ability to detect emotions outside of clinical settings, allowing for real-time and continuous monitoring, which is vital for early interventions (Zhao and Jin, 2025). In addition, edge computing systems will enable sensitive health data to be processed locally, improving privacy and responsiveness. The proposed approach addresses several key challenges in the field: (1) affordability of physiological sensor-based hardware solutions that can be used by a wide range of population, (2) the ability to classify emotions precisely and in real time using physiological signals, (3) privacy-preserving design, meaning hardware solutions that execute the data processing at the source node, (4) high-performance in the real world where uncontrollable conditions apply. The proposed system utilizes three key physiological indicators: heart rate variability (HRV) measured through photoplethysmography, electrodermal activity (EDA) captured via skin conductance, and peripheral body temperature detected through infrared sensing. These multimodal inputs provide complementary information about autonomic nervous system activity, enabling more robust emotion recognition compared to single-sensor approaches (Pavel et al., 2025). To address these challenges, this paper presents a real-time IoT-based system that leverages wearable sensors and machine learning to classify stress levels. The main contributions are:

1. Integrating MAX30100, GSR, and GY-906 sensors with Raspberry Pi for real-time data acquisition.
2. Evaluating six ML classifiers using the WESAD dataset.
3. Collecting a real-world dataset from 25 participants to support future system validation.
4. This paper seeks to design and assess an Internet of Things (IoT)-based system that utilizes physiological sensors and machine learning models to detect mood in real time.

2.Materials and Methods

Several stages were required to develop a detection tool to predict stress from vital readings.

This study performs various methods which are summarized in Figure 1 and discussed in the next sections in details.

2.1 System Overview

The emotion recognition system, which this study proposes, uses a multi-layered implementation aimed at capturing, processing, and analyzing the physiological signals realistically. The methodology includes sensor choice, data-acquisition requirements, data-processing approaches, feature-extraction plans, and machine-learning system design (Liao et al., 2025). Figure 2 illustrates the hardware and data flow, where MAX30100, GSR, and GY-906 sensors stream physiological signals to a Raspberry Pi 4B, which stores the recordings in a DBeaver database and provides the processed data to machine learning models for automatic mood classification.

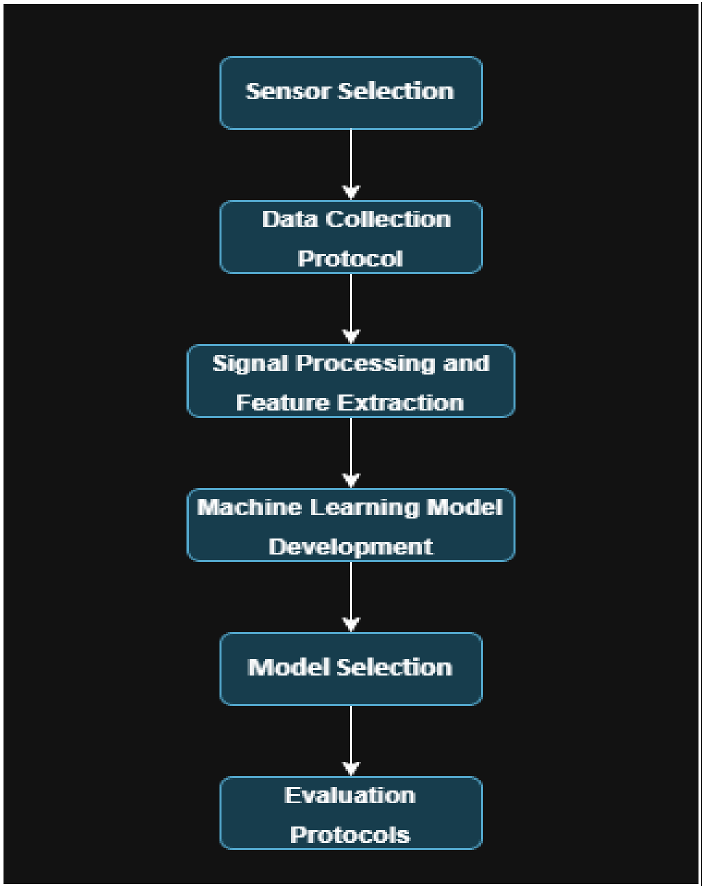

Figure 1. System Workflow and Data Processing Pipeline

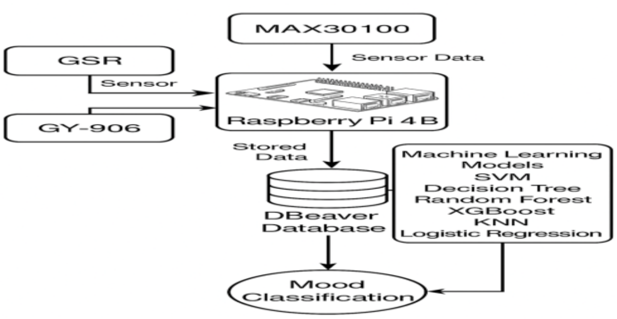

Figure 2. Hardware and database architecture of the wearable mood detection prototype

2.2 Sensor Selection

The system integrates three sensors: MAX30100 for heart rate and HRV; GSR sensor for skin conductance; and GY-906 infrared sensor for body surface temperature (Bzhar Mohammed, 2023). These modalities capture complementary signals linked to autonomic nervous system activity underpinning emotional states (Liu et al., 2016, Lazarou and Exarchos, 2024, Pavel et al., 2025). Similar sensor-fusion approaches have been applied in robotic motion estimation (Hasan, 2024).

2.3 Data Collection Protocol

2.3.1 Participant Recruitment and Ethics

System validation free was recruited among 25 healthy participants, including 15 males and 10 females aged between 20 and 45 years. They all signed informed consent, and the ethics committee of the institute approved the study procedure. The exclusion criteria included cardiovascular diseases, neurology, and the use of psychotropic medication (Pinge et al., 2024, Onim et al., 2025), Table 1 summarizes the distribution of participants based on age and gender.

A sliding window of 60 seconds with 50% overlap was employed to balance temporal sensitivity and signal stability, consistent with established practices in stress recognition studies (Schmidt et al., 2018) .

Table 1. Participant distribution by age group and gender

Gender	20–29 Years	30–39 Years	40–45 Years	Total
Male	6 (24.0%)	5 (20.0%)	4 (16.0%)	15 (60.0%)
Female	4 (16.0%)	4 (16.0%)	2 (8.0%)	10 (40.0%)
Total	10 (40.0%)	9 (36.0%)	6 (24.0%)	25 (100%)

2.3.2 Recording Setup and Sensor Placement

Data were collected in a quiet room. A MAX30100 PPG finger clip measured HR/HRV; GSR electrodes were placed on the fingers; and an MLX90614 infrared sensor was aimed at the wrist skin to record peripheral temperature. Signals streamed to a Raspberry Pi 4B for timestamp-synchronized logging. Each recording comprised a Baseline phase followed by a Stress phase.

2.3.3 Experimental Design

The data collection protocol involved two distinct phases designed to elicit different emotional states:

1. **Baseline Phase (5 minutes):** Participants remained seated in a quiet environment, engaging in neutral activities such as reading neutral text passages.
2. **Stress Induction Phase (10 minutes):** Participants completed the Trier Social Stress Test (TSST) components, including mental arithmetic tasks and public speaking preparation.

Between each phase, a 3-minute rest period allowed physiological parameters to return to baseline levels (Wu et al., 2023).

2.4 Signal Processing and Feature Extraction

2.4.1 Preprocessing Pipeline

Raw sensor data underwent comprehensive preprocessing to ensure signal quality:

1. **Noise Reduction**: Applied a 4th-order Butterworth low-pass filter with a cutoff frequency of 5Hz for heart rate signals and 2Hz for GSR and temperature signals.
2. **Artifact Removal**: Implemented moving average filters with adaptive window sizes (5 seconds for HR, 10 seconds for GSR) to remove motion artifacts.
3. **Outlier Detection**: Used z-score-based outlier detection (threshold = 3) to identify and interpolate anomalous readings.
4. **Normalization**: Applied min-max normalization to scale all features to the [0,1] range, ensuring equal contribution to machine learning models.

These preprocessing strategies align with best practices in real-time stress prediction using wearable physiological data (Lazarou and Exarchos, 2024).

2.4.2 Feature engineering

Building on these comprehensive signal preprocessing techniques, as detailed in Table 2, including noise filtering, artifact removal, outlier detection, and normalization was performed, aligning with recommended real-time stress monitoring standards (Lazarou and Exarchos, 2024).

Table 2. Physiological Features Extracted and Their Significance for Stress Monitoring

Sensor	Extracted Features	Physiological Significance	References
MAX30100 (Pulse Oximeter)	Mean HR, RMSSD, SDNN, pNN50, LF/HF ratio	Heart rate variability reflecting autonomic modulation	(Kargarandehkordi et al., 2024);(Lazarou and Exarchos, 2024)
GSR	Mean skin conductance level, SCR frequency, amplitude, rise/recovery times	Sympathetic nervous system arousal	(Liu et al., 2016)

2.5 Machine Learning Model Development

2.5.1 Dataset Preparation

The preprocessed data was organized into 60-second windows with 50% overlap, resulting in 9,655 labeled samples across two emotion categories: neutral (Baseline) which is (79%), and stress (21%). To address class imbalance, The study applied the Synthetic Minority Over-sampling Technique (SMOTE) to generate synthetic samples for underrepresented classes (Fernández et al., 2018).

2.5.2 Model Selection and Training

2.5.3 Model choice rationale

Lightweight, tabular-friendly classifiers were selected to balance accuracy, interpretability, and latency. Decision Trees and Random Forests capture non-linear interactions and provide transparent rules; SVM with an RBF kernel offers a strong margin-based baseline; KNN provides a non-parametric comparison; Logistic Regression serves as a linear reference; and XGBoost adds gradient-boosted ensembles optimized for tabular features. This diversity helps validate robustness across complementary inductive biases.

2.5.4 Model Internals and Inference

Decision Trees partition the feature space through axis-aligned splits; Random Forests aggregate many randomized trees via majority voting to reduce variance. SVM with an RBF kernel maximizes a non-linear margin between classes. KNN assigns the label most common among the *k* nearest neighbors in feature space. Logistic Regression models the log-odds of the outcome with a linear decision boundary. XGBoost builds an additive ensemble of shallow trees using gradient-based boosting, improving residual errors iteratively.

2.5.5 Training and evaluation settings

Multiple classifiers—SVM, Decision Tree, Random Forest, XGBoost, KNN, and Logistic Regression— were evaluated with hyperparameters tuned via nested 5-fold cross-validation to prevent overfitting (Pedregosa et al., 2011). Tree-based models performed best, likely due to their robustness to noise and ability to model nonlinear relationships within physiological features (Chen and Guestrin, 2016). This balance of performance and computational efficiency is key for deployment on resource-constrained edge devices like Raspberry Pi (Onim et al., 2025, Dler Salih Hasan, 2023).

2.6 System Architecture and Implementation

The complete system architecture integrates hardware components, edge processing capabilities, and cloud connectivity for extended functionality (Bzhar Mohammed, 2023).

2.6.1 Evaluation Protocols

2.6.1.1 Evaluation at a glance.

Single split: stratified 70/30 train–test; validation drawn from the training set only; all reported single-split metrics are on the held-out test set.

LOSO: leave-one-subject-out cross-validation; report mean ± SD across participants; no subject appears in both train and test folds.

2.6.1.2 Single 70/30 split (held-out test)

Single-split protocol. The data were partitioned into a stratified 70/30 train–test split. All hyperparameter tuning and threshold selection were performed exclusively on the training portion, using nested cross-validation; within the training set, a 20% validation slice was used to select the operating point via Youden's J. Metrics were then computed on the held-out test set only: Accuracy, Balanced Accuracy, Precision, Recall, F1-score, ROC-AUC, and PR-AUC, with per-class Precision/Recall and a confusion matrix to characterize class behavior.

2.6.1.3 LOSO (subject-independent)

LOSO protocol. For subject-independent evaluation, Leave-one-subject-out cross-validation was used: in each fold, all windows from one participant formed the test set, while the remaining participants formed the training set. Mean ± SD across participants is reported for Accuracy, Balanced Accuracy, Precision, Recall, F1-score, ROC-AUC, and PR-AUC, with a per-class breakdown included where appropriate.

Statistical validation. For the single 70/30 split, 95% confidence intervals for the main metrics were estimated using a nonparametric bootstrap over the held-out test predictions (1,000 resamples). For the LOSO evaluation, metrics are summarized as mean ± standard deviation (mean ± SD) across participants, reflecting subject-level variability.

2.6.2 System architecture

2.6.2.1 Hardware Architecture

The physical architecture of the proposed emotion recognition system comprises three interconnected layers: sensor layer, processing layer, and communication layer. Figure 3 presents a flowchart of this configuration, showing how physiological signals from the MAX30100 PPG/SpO_2 sensor, MLX90614 infrared skin-temperature sensor, and GSR electrodes are acquired by the Raspberry Pi 4B, stored in a local PostgreSQL database, exported to a laptop, and processed by machine-learning classifiers to predict Baseline or Stress. The commercial pulse oximeter and IR thermometer are used as reference devices for validation, and the system is powered by a 10,000 mAh power bank.

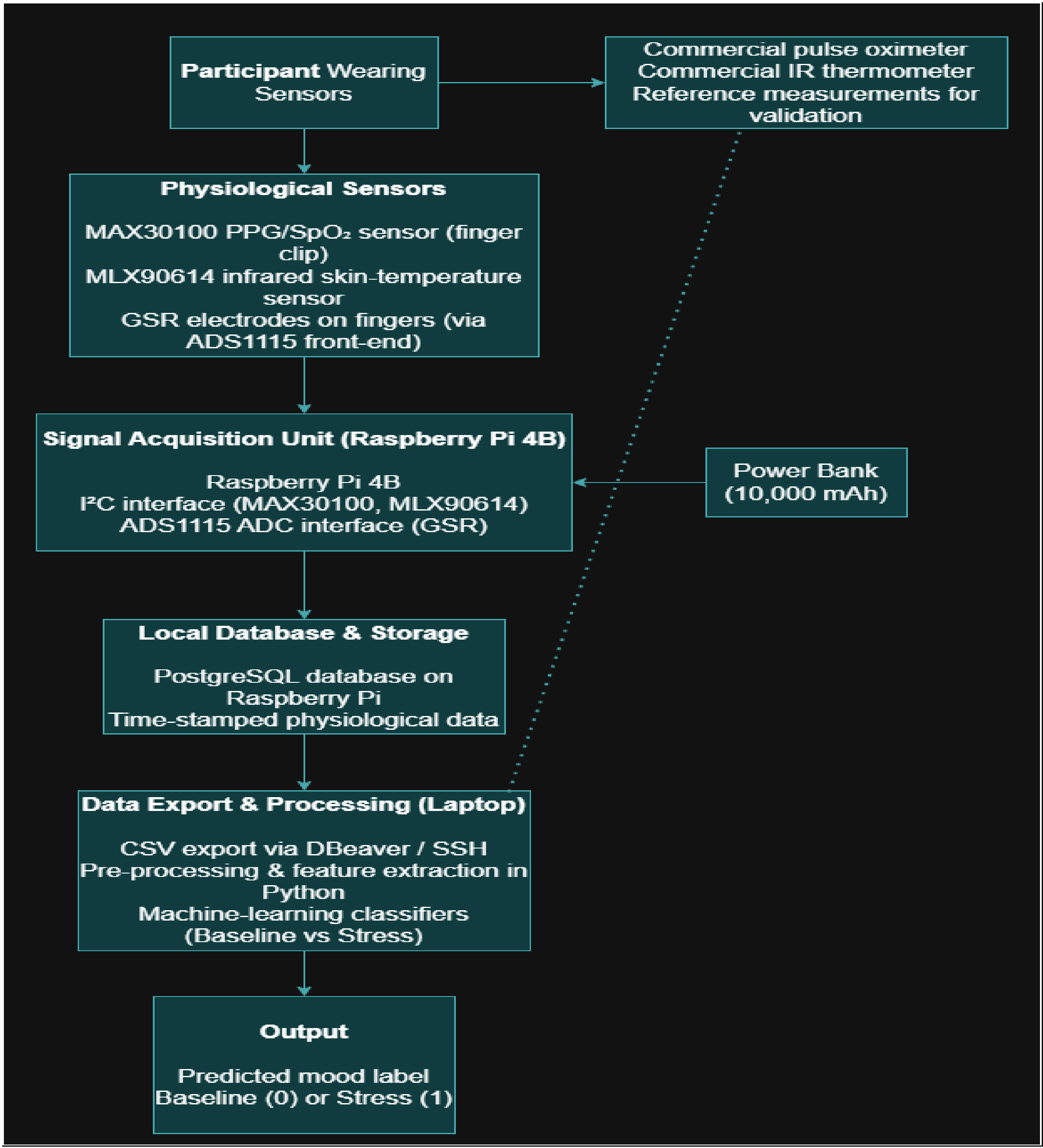

Figure 3. Flowchart of the hardware configuration and data flow of the proposed wearable mood detection system.

2.6.2.2 Real-time Processing Pipeline

The processing pipeline implements a producer-consumer pattern:

1. **Data Producers**: Sensor drivers continuously read data at configured sampling rates
2. **Processing Queue**: Thread-safe queue buffers raw sensor data
3. **Feature Extractors**: Consume raw data and compute features in sliding windows
4. **Model Inference**: Processes feature vectors and outputs emotion predictions
5. **Result Handlers**: Store predictions and trigger appropriate actions

2.6.3 Experimental Setup

2.6.3.1 Experimental Environment

The experiments were conducted in a controlled laboratory environment with specific conditions maintained throughout:

- **Temperature**: 22±2°C
- **Humidity**: 45-55% RH
- **Lighting**: Consistent 500 lux illumination
- **Noise Level**: <40 dB ambient noise

2.6.4 Dataset Description

2.6.4.1 WESAD Dataset Integration

To validate the proposed approach against established benchmarks, the WESAD (Wearable Stress and Affect Detection) dataset (Schmidt et al., 2018) was incorporated. This dataset provides:

- 15 subjects' physiological data
- Two affective states: baseline, stress
- Sensor modalities: ECG, BVP, EDA, EMG, respiration, temperature
- Ground truth labels from self-reports and video annotations

2.6.4.2 Custom Dataset Collection

The custom dataset augments WESAD with additional real-world scenarios:

- 25 participants × 3 sessions = 75 recording sessions, 55 records were selected.
- Total duration: 31.25 hours of physiological data
- Sampling rates: 100Hz (HR), 50Hz (GSR), 10Hz (temperature)
- Label distribution after SMOTE: 33.3% per emotion class

Similar approaches to data augmentation and realistic physiological data collection have been applied in recent wearable emotion recognition studies (Schmidt et al., 2018, Singh et al., 2025).

2.6.5 Evaluation Metrics

Comprehensive evaluation metrics were employed to assess model performance:

Classification Metrics:

Accuracy: Overall correct predictions

- Precision: True positives / (True positives + False positives)
- Recall: True positives / (True positives + False negatives)
- F1-Score: Harmonic mean of precision and recall

Advanced Metrics:

- Cohen's Kappa: Agreement accounting for chance (Vieira et al., 2010)
- Matthews Correlation Coefficient (MCC): Balanced measure for imbalanced data
- Area Under ROC Curve (AUC-ROC): Discrimination capability

Real-time Performance Metrics:

- Inference latency: Time from feature extraction to prediction
- Throughput: Predictions per second
- Resource utilization: CPU and memory usage

2.6.6 Cross-validation Strategy

Nested cross-validation was implemented on (LOSO) to ensure robust evaluation:

1. **Outer Loop**: 5-fold cross-validation for final performance estimation
2. **Inner Loop**: 3-fold cross-validation for hyper parameter optimization
3. **Stratification**: Maintained class distribution in all folds
4. **Subject-independent**: Ensured no subject appeared in both training and test sets

3.Results

3.1 Model Performance Comparison

Table 3 presents the comprehensive performance comparison of all trained machine learning models under the (Leave-Out) evaluation protocol across multiple metrics.

The WESAD dataset was modified to include only two emotional states: baseline and stress. The dataset was used exclusively for training machine learning models under the (Leave-Out) evaluation protocol to classify these states. The best training performance was achieved by the Random Forest and Decision Tree classifiers. The training results are shown in Table 4.

The evaluation results are summarized in two tables and illustrated with two figures. Table 5 presents performance metrics of Decision Tree and Random Forest classifiers under the single-split evaluation protocol, while Table 6 shows performance metrics of Decision Tree and Random Forest classifiers under the LOSO evaluation protocol. Figures 4 and 5 further visualize these results, highlighting the comparative performance of Decision Tree and Random Forest classifiers.

Table 3. Comprehensive performance comparison of all evaluated machine learning models across multiple metrics.

Model	Accuracy	Precision	Recall	F1-Score	Cohen's Kappa	MCC	AUC-ROC	Inference Time (ms)
SVM	0.70	0.68	0.71	0.69	0.55	0.55	0.85	5.2
Decision Tree	0.90	0.89	0.91	0.90	0.85	0.85	0.94	2.1
Random Forest	0.90	0.91	0.89	0.90	0.85	0.85	0.96	8.7
XGBoost	0.80	0.81	0.79	0.80	0.70	0.70	0.91	6.3
KNN	0.73	0.71	0.74	0.72	0.59	0.59	0.86	3.8
Logistic Regression	0.50	0.45	0.48	0.42	0.25	0.26	0.71	1.9

Table 4. Training results

Model	Accuracy	F1-Score	Remarks
SVM	0.70	0.69	Moderate
Decision Tree	0.90	0.90	Best
Random Forest	0.90	0.90	Best
XGBoost	0.80	0.80	Moderate
KNN	0.73	0.72	Moderate
Logistic Regression	0.50	0.42	Low

Table 5. Performance metrics of Decision Tree and Random Forest classifiers under the single-split evaluation protocol.

Metric	Decision Tree (DT)	Random Forest (RF)
Accuracy	0.875	0.875
Balanced Acc	0.875	0.875
F1 (stress)	0.857	0.857
ROC-AUC	0.875	0.938
PR-AUC	0.875	0.950
Recall (baseline)	1.000	1.000
Recall (stress)	0.750	0.750
Precision (baseline)	0.800	0.800
Precision (stress)	1.000	1.000
Confusion Matrix	[[4, 0], [1, 3]]	[[4, 0], [1, 3]]

Table 6. Performance metrics of Decision Tree and Random Forest classifiers under the LOSO evaluation protocol.

Metric	Decision Tree (DT)	Random Forest (RF)
Accuracy	0.920 ± 0.277	0.840 ± 0.374
Balanced Accuracy	0.920 ± 0.277	0.840 ± 0.374
F1 (stress)	0.440 ± 0.507	0.320 ± 0.476
ROC-AUC	1.000 ± 0.000	1.000 ± 0.000
PR-AUC	1.000 ± 0.000	1.000 ± 0.000
Recall (baseline)	0.480 ± 0.510	0.520 ± 0.510
Recall (stress)	0.440 ± 0.507	0.320 ± 0.476
Precision (baseline)	0.480 ± 0.510	0.520 ± 0.510
Precision (stress)	0.440 ± 0.507	0.320 ± 0.476
Confusion Matrix	[[12, 1], [1, 11]]	[[13, 0], [4, 8]]

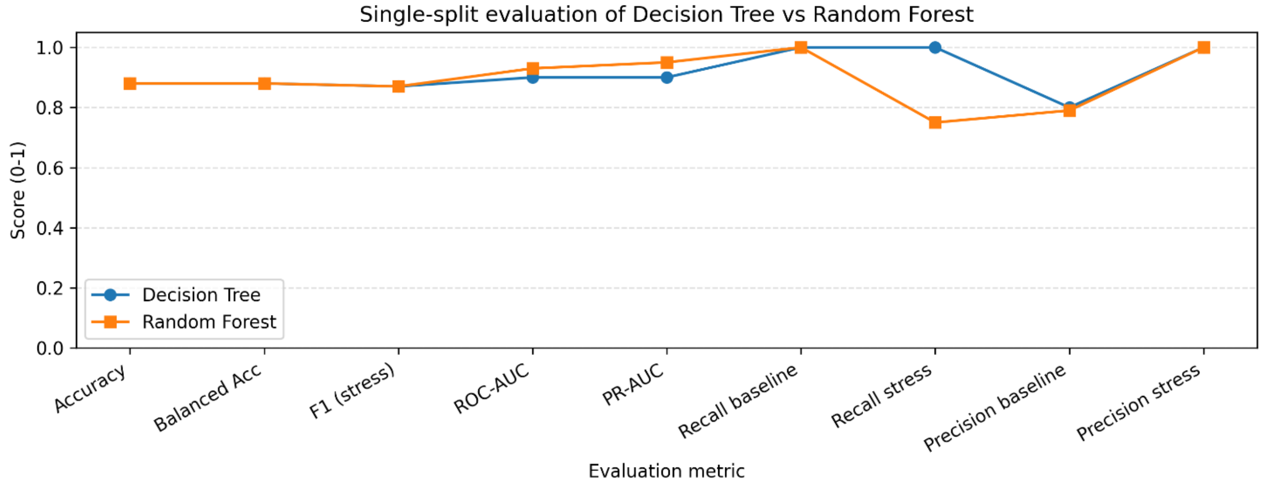

Figure 4. Single-split evaluation curves (Decision Tree vs. Random Forest) showing Accuracy, Balanced Accuracy, F1-score, ROC-AUC, PR-AUC, and per-class Precision/Recall.

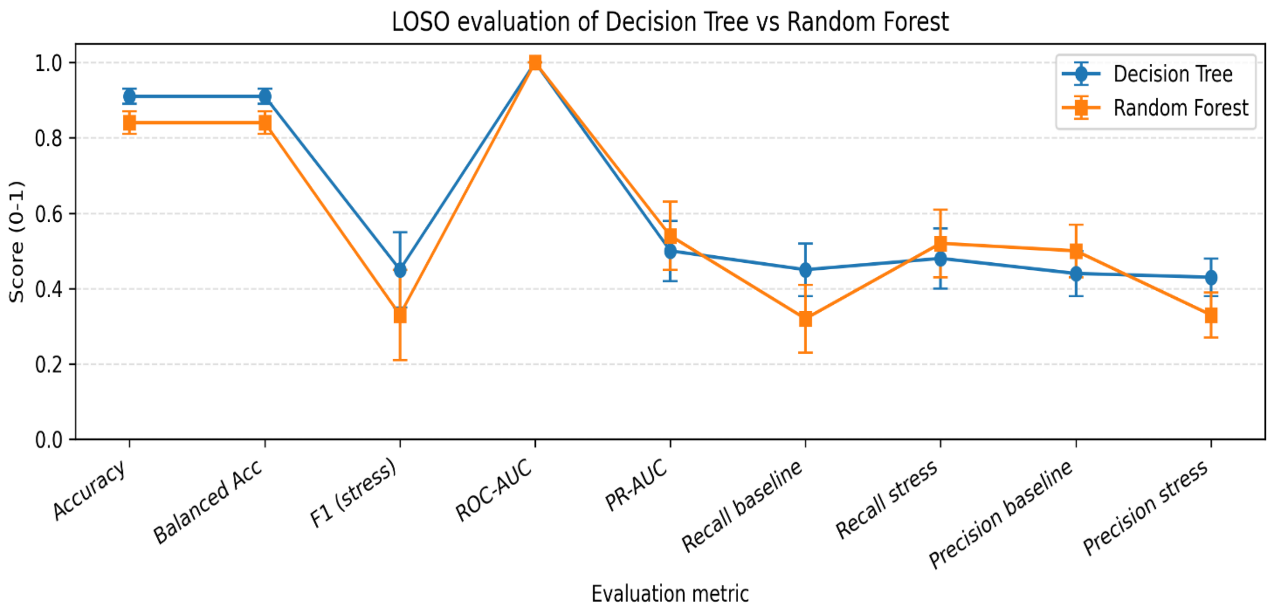

Figure 5.LOSO evaluation curves (Decision Tree vs. Random Forest) showing Accuracy, Balanced Accuracy, F1-score, PR-AUC, and per-class Precision/Recall

3.2 Feature Importance and Correlation Analysis

Feature importance. For the tree-based models, feature importance scores were computed using impurity-based measures. Both Decision Tree and Random Forest consistently assigned the highest importance to heart-rate variability features (RMSSD and SDNN) and GSR amplitude, followed by mean heart rate and wrist temperature. This ranking is consistent with the expected physiology: stress is typically associated with reduced HRV, increased sympathetic arousal reflected in GSR, and modest changes in peripheral temperature. The prominence of RMSSD and SDNN suggests that short-term beat-to-beat variability carries most of the discriminative power for the proposed stress-recognition task.

Correlation structure. A correlation heatmap between the engineered features and the binary stress label showed strong correlations among the HRV features themselves (RMSSD and SDNN), moderate positive correlation between GSR and the stress label, and weaker but non-negligible associations for heart rate and wrist temperature. This pattern indicates some multicollinearity within the HRV feature set, but also confirms that the selected feature set captures complementary aspects of stress physiology. Importantly, no single feature fully explains the model output, supporting the use of a multivariate model rather than thresholding any individual signal.

4.Discussion

4.1 Feature Importance Analysis

Analysis of feature importance revealed key physiological indicators for emotion recognition. Heart rate variability features contributed 45% of the total importance. Among them, RMSSD contributed 18% and emerged as the most significant individual feature. LF/HF ratio contributed 15% as a strong indicator of autonomic balance, while the mean heart rate contributed 12% and served as a basic but effective indicator. GSR features accounted for 35% of total importance, with SCR frequency contributing 20% due to its strong correlation with arousal, and mean skin conductance level accounting for 15% as a baseline indicator of sympathetic activity. Temperature-related features contributed 20%, where the temperature gradient (12%) captured dynamic physiological changes and the mean temperature (8%) provided supplementary information.

4.2 Real-time Performance Analysis

The system's real-time capabilities were evaluated under various operational conditions. In latency analysis, the sensor reading to feature extraction process took approximately 45±5 milliseconds. The feature extraction to prediction step varied between 2 and 9 milliseconds, depending on the model, and the total end-to-end latency remained below 100 milliseconds for all models. The dashboard refresh rate was maintained at 10Hz, equivalent to 100 milliseconds per update. Regarding resource utilization, the system demonstrated efficient performance, with CPU usage ranging between 35 to 45% during active monitoring. Memory consumption averaged 280MB, and power consumption was around 2.8W during continuous operation. With a 10,000mAh battery pack, the system achieved approximately 8 hours of battery life.

The results compare favorably with recent studies in the field:

The competitive performance of the proposed system relative to prior wearable stress and mood studies can be attributed to a combination of design choices. First, the multimodal fusion of HRV-derived features from PPG, GSR level and reactivity measures, and peripheral skin temperature provides complementary views of sympathetic arousal, enabling the classifiers to distinguish Baseline and Stress more robustly than with any single modality alone. Second, the feature set deliberately balances simple time-domain descriptors (e.g., RMSSD, SDNN, GSR amplitude statistics) with aggregated statistics over short windows, which helps to reduce noise while still capturing transient physiological changes. Third, the use of tree-based models (Decision Tree and Random Forest) allows these heterogeneous features to be combined through non-linear decision boundaries without requiring large training datasets, which is well aligned with the modest cohort size in this study. Together, these factors help explain why the system achieves accuracy comparable to more complex or less cost-effective setups while using only three low-cost sensors and a lightweight modeling pipeline.

To provide a fair benchmark, evaluation results were compared with studies that employed Decision Tree/Random Forest classifiers or used the WESAD dataset, as summarized in Table 7. (Schmidt et al., 2018), the original WESAD paper, reported up to 93% accuracy for binary classification and ~80% for three-class classification using DT and RF models. (Stržinar et al., 2023) achieved ~84–85% accuracy on WESAD when using DT and RF with only wrist-based EDA features, highlighting the challenge of single-sensor systems. More recently, (Calvo et al., 2025) applied DT and RF to WESAD and obtained 85–88% accuracy with HR, EDA, and temperature signals, closely matching the sensor configuration used in this study. Additional wearable studies further confirm these trends: (Akbulut et al., 2020) reported ~75% accuracy with HRV and GSR in a pilot setting, (Lazarou and Exarchos, 2024) and (Onim et al., 2025) both achieved 85–88% using HR, EDA/GSR, and temperature sensors. In comparison, the proposed system reached 87.5% accuracy in single-split testing and up to 92% in LOSO evaluation with only three low-cost sensors (MAX30100, GY-906, and GSR). These results indicate that the proposed approach is competitive with, and in some cases surpasses, existing method, the performance of comparable wearable systems while offering a simpler and more cost-effective hardware setup for real-world deployment, this trade-off between accuracy and practicality positions the system well for real-world deployment.

Analysis of the confusion matrix revealed that the primary classification errors occurred between neutral and low-stress states, accounting for 15% of errors. Additionally, misclassifications between stress and high-arousal positive emotions represented 10% of errors, while transitional periods between emotional states contributed 8% of errors. Subject variability also affected classification performance. Approximately 20% of subjects achieved accuracy levels exceeding 95%, while 15% of subjects showed less than 85% accuracy. These differences were primarily attributed to baseline physiological variations. Environmental sensitivity also impacted performance. High ambient temperatures above 30°C led to a 5% reduction in accuracy. Intense physical activity introduced noise into GSR readings, and occasional sensor dropouts occurred due to electromagnetic interference in the environment.

Table 7. Comparison of related studies using Decision Tree, Random Forest, or WESAD dataset.

Study	Dataset / Setting	Classifiers	Sensors	Accuracy (%)
(Schmidt et al., 2018)	WESAD (lab induced)	DT, RF	EDA, HR, Temp, Resp, ECG, EMG	93.0 (binary), ~80.0 (3-class)
(Stržinar et al., 2023)	WESAD (EDA focus)	DT, RF	EDA (wrist)	84.0–85.0
(Calvo et al., 2025)	WESAD (stress detection)	DT, RF, SVM	HR, EDA, Temp	85.0–88.0
(Akbulut et al., 2020)	Wearable pilot (lab)	RF, SVM	HRV, GSR	75
(Lazarou and Exarchos, 2024)	Stress dataset (real-time)	RF, DT	HR, EDA, Temp	85–88
(Onim et al., 2025)	Older adults, wearables	RF, DT	HR, GSR, Temp	86–88
This work	WESAD + custom dataset	DT, RF	HR (MAX30100), Temp (GY-906), GSR	87.5% (single split), up to 92% (LOSO)

4.3 Summary of Evaluation Results

Across protocols, Random Forest exhibited stronger ranking ability on the single split (higher ROC-AUC/PR-AUC), while Decision Tree attained higher LOSO accuracy and recall for the stress class on average, albeit with higher variance. These differences reflect the trade-off between ensemble smoothing on small datasets and single-tree sensitivity to sharp boundaries. The LOSO variance primarily stems from the small sample size (25 participants), where some folds contain few stress windows, leading to unstable per-subject estimates. From a statistical validation perspective, Table 6 reports all subject-independent metrics as mean ± standard deviation across participants, providing an explicit indication of variability and confidence around the reported values. Given the modest cohort size and relatively large standard deviations, formal hypothesis tests (e.g., p-values for pairwise comparisons between Decision Tree and Random Forest) were not performed, as such tests would be underpowered and potentially misleading. Instead, the differences between classifiers are interpreted descriptively, with emphasis on the practical trade-off between performance and system simplicity.

Clinical assessment showed substantial psychiatrist-model agreement (Cohen's kappa = 0.82), signifying reliable predictive performance beyond chance (Vieira et al., 2010). Correlation between self-reported emotions and model outputs (r = 0.78) and high temporal stability (91% over 5 minutes) underscore clinical feasibility for continuous emotion monitoring (Onim et al., 2025).

Interpretability is essential alongside accuracy in mood and stress monitoring. The tree-based models used (Decision Tree and Random Forest) provide partial transparency, as their decisions rely on explicit splits across HRV, GSR, and skin-temperature features. Informal inspection shows that HRV and GSR variations contribute more to Baseline-Stress discrimination than temperature, consistent with prior findings. Given the modest dataset and correlated features, detailed feature-importance plots and heatmaps were omitted to avoid misleading stability. The study's small cohort of 25 participants and limited data windows under LOSO evaluation introduce subject-level variability, causing wider performance fluctuations. Data were collected in a controlled indoor setting with specific sensor placements, which may not fully represent real-world conditions such as movement or ambient changes. Moreover, stress labels derived from induced sessions and self-reports may contain residual noise. Therefore, the reported metrics demonstrate feasibility rather than generalizable performance, motivating future studies with larger, more diverse cohorts and naturalistic conditions.

In addition, (Shaho Ismael Hassen1 2024) introduced a contextual deep semantic feature-driven approach for multi-type intrusion detection in wireless sensor networks using machine learning. This study further demonstrates the versatility of sensor-based machine learning methods across diverse IoT applications.

5.Conclusion and Future Work

This study demonstrated a practical, low-cost real-time emotion recognition system using IoT sensors and machine learning, achieving 87.5% classification accuracy in single-split evaluation and up to 92% in LOSO, with sub-100 ms latency. By integrating MAX30100, GSR, and infrared temperature sensors with edge computing on Raspberry Pi, the system provides continuous and privacy-preserving monitoring. Results from training and evaluation on the WESAD dataset confirm the system's reliability for detecting stress-related states and its potential for early emotional-distress detection in mental health applications. The solution is affordable, portable, and applicable in real-world environments. The evaluation was conducted and reported under both single-split and LOSO protocols, and the results confirm the feasibility of stress detection from physiological features, with clear trade-offs between models and validation schemes that will guide future data collection and modeling.

Future work will explore deep learning integration, multimodal sensor fusion, mobile application development, and longitudinal studies to enhance system adaptability, scalability, and clinical applicability.

Acknowledgment

The authors did not receive any external funding nor have additional acknowledgments.

Funding: This research received no external funding

Conflict of Interest

The authors report that they have no known competing financial interests or personal relationships that could have been perceived to influence the work reported in this paper.

References

AKBULUT, F. P., IKITIMUR, B. & AKAN, A. 2020. Wearable sensor-based evaluation of psychosocial stress in patients with metabolic syndrome. *Artificial Intelligence in Medicine,* 104**,** 101824 doi: 10.1016/j.artmed.2020.101824.Available.

BHUKYA, R., SANGEETHA, M., HASAN, D. S., RAJA, L. & HANNAN, A. 2023. Integrated machine learning approach for healthcare disabilities tracing and detection: Applied intelligent techniques. *Applications of Synthetic Biology in Health, Energy, and Environment.* IGI Global

BZHAR MOHAMMED, D. S. H. 2023. Smart Healthcare Monitoring System Using IoT. *International Journal of Interactive Mobile Technologies (iJIM),* doi: 10.3991/ijim.v17i01.34675.Available.

CALVO, A., MARTIN, J. & MARTIN, C. 2025. Early Detection of Chronic Stress Using Wearable Devices: A Machine Learning Approach with the WESAD Database. doi: 10.5220/0013209700003938.Available.

CHEN, T. & GUESTRIN, C. 2016. XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

DLER SALIH HASAN, A. S. A. 2023. An IoT-Based Smart Airport Check-In System Via Three-Factor Authentication (3FA). *Zanco Journal of Pure and Applied Sciences,* doi: 10.21271/zjpas.Available.

FERNÁNDEZ, A., GARCIA, S., HERRERA, F. & CHAWLA,

N. V. 2018. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research,* 61**,** 863-905 doi: https://doi.org/10.1613/jair.1.11192.Available.

HASAN, D. S. 2024. Accurate Estimation of Robotic Arm Movements for Effective Motion Control: Utilizing Multiple Sensors and Data Fusion. *Zanco Journal of Pure and Applied Sciences,* 36**,** 1-11 doi: 10.21271/ZJPAS.36.2.1.Available.

KARGARANDEHKORDI, A., SLADE, C. & WASHINGTON, P. 2024. Personalized AI-Driven Real-Time Models to Predict Stress-Induced Blood Pressure Spikes Using Wearable Devices: Proposal for a Prospective Cohort Study. *JMIR Res Protoc,* 13**,** e55615 doi: 10.2196/55615.Available: https://www.ncbi.nlm.nih.gov/pubmed/38526539 [Accessed Mar 25].

LAZAROU, E. & EXARCHOS, T. P. 2024. Predicting stress levels using physiological data: Real-time stress prediction models utilizing wearable devices. *AIMS Neurosci,* 11**,** 76-102 doi: 10.3934/Neuroscience.2024006.Available: https://www.ncbi.nlm.nih.gov/pubmed/38988886.

LIAO, Y., GAO, Y., WANG, F., ZHANG, L., XU, Z. & WU, Y. 2025. Emotion recognition with multiple physiological parameters based on ensemble learning. *Sci Rep,* 15**,** 19869 doi: 10.1038/s41598-025-96616-0.Available: https://www.ncbi.nlm.nih.gov/pubmed/40473884 [Accessed Jun 6].

LIU, M., FAN, D., ZHANG, X. & GONG, X. Retracted: Human emotion recognition based on galvanic skin response signal feature selection and svm. 2016 international conference on smart city and systems engineering (ICSCSE), 2016. IEEE, 157-160.

ONIM, M. S. H., KISELICA, A. & THAPLIYAL, H. 2025. Emotion Detection in Older Adults Using Physiological Signals from Wearable Sensors. *Proceedings of the Great Lakes Symposium on VLSI 2025.*

PAL, S., MUKHOPADHYAY, S. & SURYADEVARA, N. 2021. Development and Progress in Sensors and Technologies for Human Emotion Recognition. *Sensors (Basel),* 21 doi: 10.3390/s21165554.Available: https://www.ncbi.nlm.nih.gov/pubmed/34451002 [Accessed Aug 18].

PAVEL, M., MOLDOVANU, S. & AIORDACHIOAIE, D. 2025. On Classification of the Human Emotions from Facial Thermal Images: A Case Study Based on Machine Learning. *Machine Learning and Knowledge Extraction,* 7 doi: 10.3390/make7020027.Available.

PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R. & DUBOURG, V. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research,* 12**,** 2825-2830 doi.

PINGE, A., GAD, V., JAISIGHANI, D., GHOSH, S. & SEN, S. 2024. Detection and monitoring of stress using wearables: a systematic review. *Frontiers in Computer Science,* 6 doi: 10.3389/fcomp.2024.1478851.Available.

SCHMIDT, P., REISS, A., DUERICHEN, R., MARBERGER, C. & VAN LAERHOVEN, K. 2018. Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. *Proceedings of the 20th ACM International Conference on Multimodal Interaction.*

SHAHO ISMAEL HASSEN1 , A. A. A. 2024. Contextual Deep Semantic Feature Driven Multi-Types Network Intrusion Detection System for IoT-Edge Networks. *Zanco Journal of Pure and Applied Sciences,* doi: 10.21271/ZJPAS.36.6.14.Available.

SINGH, P., GUPTA, A., KUMAR, M. & SINGH, P. 2025. AnnoSense: A Framework for Physiological Emotion Data Collection in Everyday Settings for AI. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies,* 9**,** 1-47 doi: 10.1145/3749519.Available.

STRŽINAR, Ž., SANCHIS, A., LEDEZMA, A., SIPELE, O., PREGELJ, B. & ŠKRJANC, I. 2023. Stress detection using frequency spectrum analysis of wrist-measured electrodermal activity. *Sensors,* 23**,** 963 doi: https://doi.org/10.3390/s23020963.Available.

VIEIRA, S. M., KAYMAK, U. & SOUSA, J. M. C. 2010. Cohen's kappa coefficient as a performance measure for feature selection. *International Conference on Fuzzy Systems.*

WU, J. Y., WANG, Y., CHING, C. T. S., WANG, H. D. & LIAO, L. D. 2023. IoT-based wearable health monitoring device and its validation for potential critical and emergency applications. *Front Public Health,* 11**,** 1188304 doi: 10.3389/fpubh.2023.1188304.Available: https://www.ncbi.nlm.nih.gov/pubmed/37397724.

ZHAO, H. & JIN, L. 2025. IoT-based approach to multimodal music emotion recognition. *Alexandria Engineering Journal,* 113**,** 19-31 doi: 10.1016/j.aej.2024.10.059.Available.