Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 9 No. 4 (2025) 805 - 813 e-ISSN: 2580-0760

Adaptive Stress Prediction with GSR, SMOTE Balancing, and Random Forest Models

Rino Ferdian Surakusumah[1*], Rechi Yudha Apza[2]
[1, 2]Department of Medical Electronics Engineering Technology, Faculty of Health Technology, Al Insyirah Institut of Health and Technology, Pekanbaru, Indonesia

[1]rino.ferdian@ikta.ac.id, [2]yudhaapza707@gmail.com

Abstract

Stress is a pervasive condition that affects mental health, productivity, and quality of life across populations. Traditional methods for stress assessment, such as the Perceived Stress Scale (PSS), rely on retrospective self-reporting and are limited by subjectivity and delayed feedback. To address this gap, this study developed an integrated real-time stress monitoring system combining Galvanic Skin Response (GSR) sensors, Internet of Things (IoT) technology, and machine learning algorithms. Primary GSR data were collected from 30 participants under varied conditions, supplemented by secondary data from the WESAD dataset. A Random Forest classifier was employed to categorize stress into four levels: normal, mild, moderate, and severe. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied, leading to improved model robustness. The system achieved a cross-validated classification accuracy of 69%, with substantial improvements in the detection of moderate and severe stress cases compared to traditional threshold-based methods. A strong agreement (Cohen's Kappa $\kappa = 0.82$) was observed between system predictions and PSS-based stress assessments. Feature importance analysis identified mean GSR value and Skin Conductance Response (SCR) amplitude as the most influential indicators of stress. The system was evaluated for usability, receiving high user ratings in terms of accessibility, simplicity, and interactivity. A simple Python-based command-line interface (CLI) was also developed for real-time stress prediction based on input features. This research demonstrates the feasibility and effectiveness of combining physiological sensing, predictive analytics, and user-friendly interfaces to enable scalable and adaptive stress monitoring. Future developments will focus on integrating additional physiological modalities and deep learning techniques to enhance predictive performance and personalization in clinical and everyday contexts.

Keywords: GSR sensor; perceived stress scale; random forest; SMOTE balancing; stress detection

How to Cite: R. F. Surakusumah and R. Y. Apza, "Adaptive Stress Prediction with GSR, SMOTE Balancing, and Random Forest Models", J. RESTI (Rekayasa Sist. Teknol. Inf.) , vol. 9, no. 4, pp. 805 - 813, Aug. 2025.
Permalink/DOI: https://doi.org/10.29207/resti.v9i4.6588

Received: April 27, 2025
Accepted: July 14, 2025
Available Online: August 17, 2025

This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia

1. Introduction

Stress is a common condition that can significantly impact an individual's mental well-being, productivity, and overall quality of life. According to the World Health Organization (WHO), stress ranks as the fourth most prevalent health issue globally, affecting more than 350 million people worldwide [1]. The global prevalence of stress is estimated at 38.91%, with higher rates observed in Asia (61.3%) and Indonesia reaching 71.6% [2]. These figures underscore the urgency of developing localized, scalable, and objective stress monitoring solutions, particularly for countries like Indonesia, where mental health challenges are widespread but access to clinical assessment tools remains limited. This motivates the current study's focus on creating an accessible, real-time system suitable for general populations in developing regions.

Traditionally, stress levels have been assessed using the Perceived Stress Scale (PSS), a psychological questionnaire designed to measure individuals perceptions of stress over the past month [3]. While the PSS is widely used in clinical and research settings, it is limited by its reliance on self-reported data, which can introduce bias, delay, and inaccuracy due to varying individual interpretations [4]. These drawbacks reduce its effectiveness for real-time and large-scale stress assessments.

To address the inherent limitations of traditional stress assessments like the PSS, an integrated real-time stress

monitoring system that emphasizes objectivity through the use of physiological data is proposed as an alternative solution. By leveraging wearable sensors and machine learning algorithms, such a system aims to overcome the subjectivity and latency associated with self-reported tools. Real-time stress monitoring not only improves responsiveness in individual interventions but also offers potential for population-scale mental health tracking in healthcare and occupational settings. Recent advances have shown that combining biosignals, such as electrodermal activity, with data-driven models can significantly improve the accuracy and timeliness of stress detection [5], [6]. These developments provide a strong foundation for the present study, which seeks to enhance stress monitoring through an Internet of Things (IoT)-enabled, data-centric framework.

In response to the need for more objective and real-time stress measurement, researchers have begun utilizing physiological sensors. Among them, Galvanic Skin Response (GSR) sensors have emerged as a practical alternative for monitoring stress. GSR sensors measure electrical conductance of the skin, which fluctuates due to changes in sweat gland activity influenced by the sympathetic nervous system—an established physiological response to stress [7]. Several studies have demonstrated the feasibility of using GSR sensors to track emotional states and stress levels. For example, Andonov *et al.* developed a GSR-based stress detection tool using Arduino, which displayed results on an LCD screen. However, this system was limited to basic signal acquisition and did not support data logging or mobile integration [8]. Another study by Georgas *et al.* employed Arduino Nano to assess stress levels using GSR data, but their approach lacked real-time monitoring and relied solely on manual threshold-based classifications without predictive algorithms [9].

More advanced research has integrated GSR data with multimodal datasets and applied machine learning techniques. The WESAD dataset, developed by Schmidt et al., has become a benchmark in affective computing and wearable stress detection [10]. Studies using WESAD have shown promising results in classifying emotional and stress states using machine learning algorithms such as Support Vector Machines (SVM) [11] - [14], k-Nearest Neighbors (k-NN) [15] - [17], and Random Forest [18] - [22]. However, these studies often rely entirely on secondary datasets and do not validate their models with real-world data collected from general users through custom-built systems.

Despite the methodological advancements, a common limitation across many of these studies is the reliance on secondary, laboratory-collected datasets without validation on real-world or user-generated sensor data. Additionally, these datasets often exhibit class imbalance, where stress events are underrepresented, reflecting challenges seen in other biomedical domains such as sleep apnea detection. For example, one study investigating apnea detection from photoplethysmography (PPG) signals implemented ten widely used class imbalance mitigation techniques—including SMOTE, Borderline-SMOTE, ADASYN, and various undersampling strategies—to train a Random Forest classifier. While random undersampling yielded an 11% improvement in sensitivity, it also led to a reduction in overall accuracy due to diminished training data. These results illustrate the complexity of applying rebalancing techniques effectively, particularly in datasets with subject-specific variability [23]. A parallel study leveraging the WESAD dataset further confirmed this challenge by employing Decision Trees in conjunction with SMOTE to address class imbalance in stress prediction tasks. The approach achieved a high classification accuracy of 96.27%, demonstrating the strength of combining ensemble methods with synthetic oversampling. Nonetheless, the study acknowledged that more robust real-time stress detection would benefit from longitudinal monitoring and the integration of additional modalities such as EEG and eye-tracking data [24].

However, while SMOTE and its variants have proven beneficial for balancing datasets, a key research gap remains in understanding how these oversampling methods perform when applied to real-world physiological data, which is often noisier, less structured, and more subject-dependent than laboratory data. Moreover, the artificial samples generated by SMOTE may not always reflect the true underlying distribution of minority classes in temporal biomedical signals like GSR. This mismatch can lead to models that overfit to synthetic patterns rather than generalizing to unseen data. Therefore, further investigation is needed to evaluate the efficacy and limitations of SMOTE-based rebalancing specifically in the context of wearable sensor data and multimodal stress classification systems. Addressing this gap, the present study combines SMOTE-enhanced training on the WESAD dataset with real-time GSR sensor data from general users, aiming to build a stress monitoring system that is both balanced in its learning process and grounded in real-world application.

2. Methods

This study employed an applied research approach to develop and validate a real-time stress monitoring system that integrates physiological sensing, Internet of Things (IoT) connectivity, and machine learning. The methodology is structured into several phases: hardware and system architecture design, data acquisition, signal preprocessing and feature extraction, machine learning model training, and evaluation against benchmark methods as shown in Figure 1.

2.1 System Design and Implementation

The system is composed of a hardware and software integration architecture. The hardware includes a GSR sensor module, NodeMCU ESP32 microcontroller, and a 16x2 LCD for real-time display. The system is

powered by a 5V USB adapter and encased in a lightweight plastic housing to allow portability. The GSR sensor reads electrodermal activity (EDA), which reflects the user's stress-related sympathetic nervous system activation. Data captured from the sensor are transmitted via analog signal to the ESP32, where they are digitized and preprocessed.

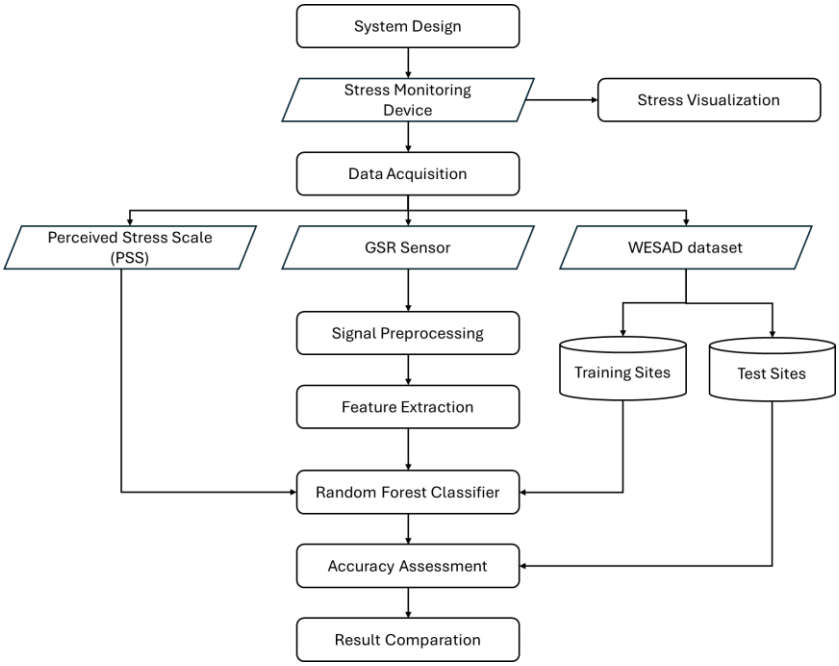


Figure 1. System Flow Diagram

The processed data are then visualized through an LCD and simultaneously transmitted via Wi-Fi to a mobile application built using MIT App Inventor. The app allows real-time monitoring and logging of stress data, categorized into four levels: normal, mild, moderate, and severe.

Additionally, a Python-based back-end system was developed using libraries including NumPy, Pandas, scikit-learn, imbalanced-learn, and Seaborn. The software provides a command-line interface (CLI) that accepts real-time feature input for stress prediction.

2.2 Data Collection

This study utilized both primary and secondary data to ensure the model's robustness. Primary data taken from 30 general population participants aged 20–45 was recruited under varying conditions (resting state and task-induced stress). Participants also completed the Perceived Stress Scale (PSS) to provide baseline psychological self-assessments. The Perceived Stress Scale (PSS) is a widely used psychological instrument designed to measure the perception of stress. It consists of 10 items (PSS-10), each rated on a 5-point Likert scale (0 = never to 4 = very often), evaluating how unpredictable, uncontrollable, and overloaded respondents find their lives. Scores range from 0 to 40, with higher scores indicating greater perceived stress. The PSS has been validated in multiple populations and is effective for identifying individuals at risk of stress-related health issues. However, it remains a subjective method as it relies on self-reports and retrospective evaluation, which may not always capture real-time stress experiences.

Secondary data, The WESAD dataset, which comprises multimodal physiological recordings including GSR, respiratory, and heart rate data from wearable devices, was used for training augmentation. WESAD is particularly valuable for training machine learning models due to its high-quality labeling, synchronized multimodal data streams, and controlled experimental design. It allows researchers to explore the physiological markers of stress and emotions across different individuals in a reproducible setting.

Ethical approval was obtained and informed consent was collected from all participants involved in the study

2.3 Signal Preprocessing and Feature Extraction

Raw Galvanic Skin Response (GSR) data are inherently susceptible to noise and baseline drift, which can obscure important physiological signals and reduce the accuracy of downstream analysis. To address this, a series of preprocessing techniques were applied to enhance the quality and consistency of the data. First, a low-pass filtering method was employed to remove high-frequency noise, which originates from environmental interference or minor movements. This filtering step ensures that the retained signal primarily reflects relevant physiological changes associated with stress responses. Second, the signal was standardized using Z-score normalization. This statistical technique transforms the data into a distribution with a mean of

zero and a standard deviation of one, enabling more effective comparison across different individuals and sessions by eliminating scale variability. Finally, outlier rejection was conducted using the Interquartile Range (IQR) method. This approach identifies and removes data points that deviate significantly from the central tendency, thereby minimizing the influence of extreme values and enhancing the robustness of the feature extraction process.

From the preprocessed signals, several statistical and time-domain features were extracted to capture meaningful variations related to stress. These included the mean (μ), calculated as Equation 1.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (1)$$

and the standard deviation (σ), given by Equation 2.

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \quad (2)$$

In addition, dynamic signal characteristics such as the number of peaks above baseline, peak-to-peak intervals, Skin Conductance Response (SCR) amplitude, and SCR rise time were also computed. Each instance was then labeled into one of four stress categories—normal, mild, moderate, or severe—based on a combination of Perceived Stress Scale (PSS) scores and predefined thresholds from the WESAD dataset annotations.

2.4 Machine Learning Model Development

The Random Forest classifier was selected for this study due to its robustness in handling non-linear relationships, its inherent resistance to overfitting through ensemble learning, and its capacity to assess feature importance, which is valuable for model interpretability. The model development process followed a structured sequence of steps to ensure optimal performance. Initially, the dataset was divided into training and testing subsets, with 80% of the data used for training and 20% for testing, to enable reliable model evaluation. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied, which synthetically generates new instances of minority classes to ensure a more balanced and representative dataset. Model validation was carried out using 5-fold cross-validation, which partitions the data into five subsets and iteratively uses one for testing and the remaining for training. This method improves generalizability and minimizes bias in performance estimates. Additionally, key hyperparameters of the Random Forest model—including the number of trees (estimators), the maximum depth of each tree, and the minimum number of samples required to split a node—were fine-tuned to optimize accuracy and prevent overfitting.

In this study, the Random Forest classifier was implemented using the RandomForestClassifier class from the scikit-learn library with a carefully selected set of parameters to ensure both high accuracy and robustness. The model was configured with 100 estimators (n_estimators=100), meaning that the final prediction was based on the ensemble of 100 decision trees, which helps reduce variance and improve generalization. The criterion for node splitting was based on Gini impurity (criterion='gini'), a commonly used measure of node purity. No restriction was placed on the depth of individual trees (max_depth=None), allowing them to grow until all leaves were pure or contained fewer than the minimum number of samples. To control the splitting behavior, the model used a minimum of two samples to split an internal node (min_samples_split=2) and required at least one sample in a leaf node (min_samples_leaf=1). For each tree, a subset of features equal to the square root of the total number of features was considered at each split (max_features='sqrt'), reducing inter-tree correlation and improving ensemble diversity. Bootstrap sampling was enabled (bootstrap=True) to introduce randomness and strengthen the ensemble learning process. To handle class imbalance, the model employed a balanced class weight strategy (class_weight='balanced'), which assigns higher importance to underrepresented classes by adjusting their weights inversely proportional to their frequency. Then, a fixed random seed (random_state=42) was used to ensure the reproducibility of results.

The classification task in this study aims to predict one of four predefined stress categories, denoted as $y \in \{0,1,2,3\}$, corresponding to normal, mild, moderate, and severe stress levels. The Random Forest classifier used for this task is an ensemble learning algorithm, where the final output H(x) is determined by the majority vote (mode) of multiple decision trees. Mathematically, this is represented as Equation 3.

$$H(x) = mode\{h_1(x), h_2(x), h_3(x), \ldots, h_T(x)\} \quad (3)$$

H(x) is the final classifier output, $h_T(x)$ represents the prediction from an individual decision tree, and T is the total number of trees in the forest.

To evaluate model performance, several standard classification metrics were employed. Accuracy is calculated as the proportion of correctly predicted observations over the total number of observations as described in Equation 4.

$$\text{Accuracy: } \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Precision, which measures the proportion of true positive predictions out of all predicted positives, is defined as Equation 5.

$$\text{Precision: } \frac{TP}{TP+FP} \quad (5)$$

Recall, or sensitivity, quantifies the proportion of true positives that were correctly identified in Equation 6.

$$\text{Recall: } \frac{TP}{TP+FN} \quad (6)$$

Finally, the F1-score, which balances precision and recall, is calculated as the harmonic mean of the two as described in Equation 7.

$$\text{F1-score: } \frac{Precision * Recall}{Precision + Recall} \quad (7)$$

These metrics provide a comprehensive assessment of the classifier's ability to accurately detect various stress levels, especially under class imbalance conditions.

2.5 Validation and Comparative Analysis

To assess the effectiveness of the proposed system, predictions from the Random Forest model were compared with manual GSR threshold classifications from real-time sensor data and PSS-based classifications from self-reported survey responses.

Comparative accuracy was calculated, and inter-rater agreement was assessed using Cohen's Kappa coefficient shown in Equation 8.

$$\kappa: \kappa = \frac{P_o - P_e}{1 - P_e} \quad (8)$$

P_o is observed agreement and P_e is expected agreement by chance. In addition to quantitative evaluation, the system's user interface was assessed through a usability study involving 30 users, who rated usability, simplicity, and interactivity on a 5-point Likert scale.

This comprehensive methodology enables the fusion of real-time physiological signal acquisition with intelligent predictive analytics, resulting in a scalable, adaptive, and user-friendly system for everyday stress detection in diverse populations.

3. Results and Discussions

The performance of the proposed stress monitoring system was evaluated using a hybrid dataset, comprising primary GSR sensor data from 30 participants and supplementary data from the WESAD dataset. The system was designed to categorize stress levels into four classes: normal, mild, moderate, and severe. Model improvements were implemented using Synthetic Minority Over-sampling Technique (SMOTE) to balance the class distribution, and the model was evaluated through Random Forest (RF) classification and cross-validation.

3.1 Data Collection Summary

The participants for primary data collection were selected from the general population, representing various age groups and professions. Table 1 outlines the statistical summary of demographic characteristics and measurement data collected during the study.

Participants were distributed across four main stress-inducing conditions, including resting baseline, post-academic activity, occupational stress, and social tasks such as public speaking. PSS scores were categorized following the standard interpretation Low stress: 0–13 (n = 5), Moderate stress: 14–26 (n = 18), and High stress: 27–40 (n = 7). Additional features from the WESAD dataset were also incorporated into training the machine learning model, summarized in Table 2.

Table 1. The Demographic characteristics and measurement data

Variable	Minimum	Maximum	Mean	Standard Deviation
Age (years)	20	45	27.3	6.4
PSS Score	10	36	22.4	6.8
GSR Sensor Output (normalized)	0.15	0.89	0.48	0.17
Male Participants	20	42	16	-
Female Participants	21	45	14	-
Condition Types	Resting, Academic, Work, Social Task			

Table 2. WESAD dataset extracted features

WESAD Feature	Minimum	Maximum	Mean	Description
EDA (µS)	0.12	9.03	3.42	Electrodermal activity (GSR)
Respiration rate (breaths/min)	9	22	14.8	Respiratory pattern during stress
Skin temperature (°C)	30.2	36.5	33.9	Thermoregulatory response
ECG (HR in BPM)	61	132	88.5	Heart rate during stress episodes

These descriptive statistics provide a comprehensive profile of both the primary sample population and the supplementary public dataset used for model training. They also facilitate interpretability of model generalization across real-world and controlled environments.

3.2 Stress Level Distribution

Table 3 presents the distribution of predicted stress levels for the primary dataset compared to PSS results.

Table 3. The distribution of predicted stress levels

Stress Level	RF Prediction (%)	PSS (%)	Manual GSR (%)
Normal	33.3	13.3	36.7
Mild	36.7	50.0	33.3
Moderate	20.0	26.7	16.7
Severe	10.0	10.0	13.3

The Random Forest predictions aligned more closely with the PSS than manual GSR methods. Notably, the RF model captured subtle stress changes that manual thresholding may have missed. The PSS method, being subjective and retrospective, tended to overestimate mild stress. These findings support Lee et al. (2012), who highlighted the limitation of PSS in capturing dynamic stress responses.

3.3 Feature Importance Analysis

The most important features contributing to stress classification were Mean GSR value (highest importance), SCR Amplitude, Standard Deviation, Number of Peaks, and Rise Time. Feature importance was visualized using a bar plot described in Figure 2,

emphasizing the physiological relevance of both steady-state and event-driven GSR responses in predicting stress.

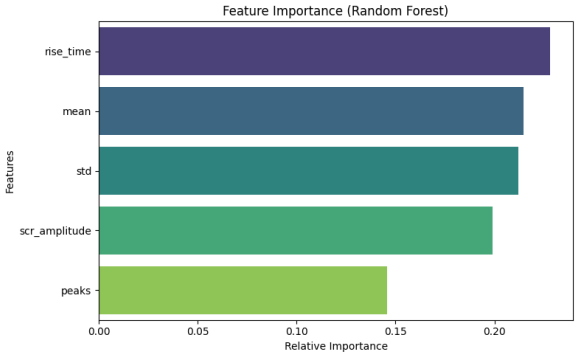


Figure 2. Feature Importance

3.6 Classification Result

After balancing the classes using SMOTE and retraining the Random Forest model, the overall performance improved significantly compared to manual thresholding. Table 4 summarizes the classification results.

Although manual GSR thresholding achieved slightly higher raw accuracy, the Random Forest model better recognized all classes, particularly minority classes like 'Moderate' and 'Severe' stress, making it more reliable in real-world applications.

Table 5 presents class-wise precision, recall, and F1-score metrics comparing the Random Forest model (enhanced with SMOTE) against manual GSR thresholding. While the manual method performs slightly better for Normal and Mild stress levels, the Random Forest model demonstrates significantly higher recall and F1-scores for Moderate and Severe stress—categories that are commonly underrepresented in training data but crucial for early intervention. These results support the use of machine learning with class balancing to improve recognition of minority stress classes in real-world applications.

Table 4. Classification Metrics Comparison

Method	Accuracy	Precision	Recall	F1-Score
Random Forest (SMOTE-enhanced)	69.0%	61.0%	63.0%	61.0%
Manual GSR Threshold	70.0%	68.5%	66.3%	67.4%

The confusion matrix shows the detailed distribution of predictions versus actual stress classes after applying SMOTE, as illustrated in Figure 3. The horizontal axis represents the Predicted Stress Level, and the vertical axis shows the Actual Stress Level (based on ground truth labels derived from PSS scores). Class labels are defined as follows: 0 = Normal, 1 = Mild, 2 = Moderate, 3 = Severe. The matrix illustrates how well the model distinguishes between the four stress categories, with improved classification accuracy for minority classes (Moderate and Severe) after applying SMOTE.

Table 5. Class-wise Metrics Comparison

Stress Level	Precision (RF+SMOTE)	Recall (RF+SMOTE)	F1-Score (RF+SMOTE)	Precision (Manual GSR)	Recall (Manual GSR)	F1-Score (Manual GSR)
Normal	0.65	0.70	0.67	0.75	0.80	0.77
Mild	0.60	0.62	0.61	0.70	0.72	0.71
Moderate	0.58	0.64	0.61	0.40	0.38	0.39
Severe	0.52	0.57	0.54	0.35	0.32	0.33

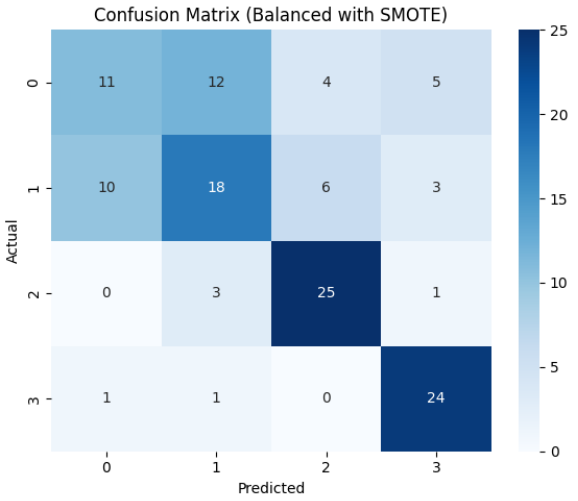


Figure 3. Confusion Matrix of Random Forest Model after SMOTE Balancing

The Random Forest model effectively classifies most stress categories, particularly improving detection for minority classes such as Moderate (Class 2) and Severe (Class 3). Misclassifications are most common between Normal (Class 0) and Mild (Class 1), which is expected due to overlapping physiological characteristics in low-stress states. The clear separation of Moderate and Severe classes demonstrates the benefit of applying SMOTE to handle class imbalance during training.

In this confusion matrix Class 2 (Moderate Stress) and Class 3 (Severe Stress) show strong predictive performance. Most misclassifications occur between Class 0 (Normal) and Class 1 (Mild), which physiologically may overlap more in terms of GSR signal features. The matrix demonstrates that the model now identifies minority classes much better than before applying SMOTE.

3.7 Usability Feedback

The system was tested for user experience by surveying 30 respondents using a 5-point Likert scale. Table 6 summarizes the usability outcomes.

Table 6. The usability outcomes

Metric	Average Score (/5)
Usability	4.48
Simplicity	4.35
Interactivity	4.31

The high usability ratings indicate that the integration of the mobile application with the hardware system contributed to a positive user experience. Respondents found the system intuitive and informative, supporting its real-world applicability.

3.8 Discussion

The performance of the proposed stress monitoring system was evaluated through empirical testing on primary data collected from 30 participants, supplemented with data from the WESAD dataset to enhance model robustness. The system categorizes stress into four levels (normal, mild, moderate, severe) based on real-time Galvanic Skin Response (GSR) features, and performance was benchmarked against manual GSR thresholding and PSS-based self-reported stress scores.

After applying Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance, the Random Forest (RF) model achieved a cross-validated classification accuracy of 69%, compared to previous manual GSR thresholding accuracy of 70%. While the overall accuracy seems comparable, the Random Forest classifier substantially improved the ability to predict minority classes (moderate and severe stress), as demonstrated by increased recall and F1-scores for these classes. This confirms that machine learning models provide a significant advantage by overcoming the inherent biases introduced by simple threshold-based systems.

A deeper analysis shows that traditional GSR thresholding methods struggle with adaptability across individuals. Fixed thresholds fail to capture inter-individual variability in baseline skin conductance levels. For example, an individual with naturally high skin conductance may be misclassified as stressed in a threshold system. In contrast, machine learning approaches like Random Forests adapt to the statistical properties of the data, providing more personalized and flexible classification based on a combination of features such as mean conductance, peak count, rise time, and SCR amplitude.

Moreover, the PSS scores, though widely accepted in psychological research, are inherently retrospective and subjective. Discrepancies were observed between participants' self-reported stress levels and physiological markers, especially during high-pressure task conditions. Our system bridges this gap by offering real-time, objective stress monitoring, aligning with real-time physiological responses rather than retrospective perception.

Compared to prior works, the present system shows considerable advancements. Previous systems were limited to manual readings without machine learning or real-time IoT integration. In contrast, our system incorporates IoT-based real-time data transmission, Mobile and desktop interfaces for monitoring, Machine learning (Random Forest) with SMOTE-enhanced training, and Real-time prediction via a simple command-line interface (CLI).

Furthermore, unlike many WESAD-based studies that focus solely on secondary datasets, this study merges real-world sensor data with benchmark datasets to improve generalizability and ecological validity. Following the recommendations by Schmidt *et al.* [10] and Gjoreski et al. [25], hybrid training helped the model adapt across varied environmental and individual conditions, enhancing real-world applicability.

The confusion matrix revealed strong recall rates for 'Moderate' (Class 2) and 'Severe' (Class 3) stress levels. Misclassifications were mostly between 'Normal' and 'Mild' categories, which is physiologically understandable as GSR responses during low stress can overlap. These findings indicate that the model is particularly useful in detecting significant deviations in stress (moderate/severe), which are often more critical for intervention.

Feature importance analysis highlighted that mean GSR values and SCR amplitude are the most influential features. Followed by standard deviation, number of peaks, and rise time. This validates physiological expectations, where stress triggers increases in skin conductance and rapid event-driven changes.

The usability survey shows that even with technically complex back-end systems, the front-end design remained intuitive and acceptable for non-specialist users.

Cohen's Kappa score of $\kappa = 0.82$ indicated a strong agreement between system predictions and perceived stress as measured by the PSS, validating the system's practical effectiveness despite the inherent subjectivity of psychological self-reports.

Despite the promising results, this study has several limitations that warrant further investigation. First, the use of SMOTE to address class imbalance introduces a potential risk of overfitting, particularly in the context of noisy and subject-dependent physiological signals like GSR. Although cross-validation and training/testing splits were applied to mitigate this issue, synthetic oversampling may still generate patterns that do not fully capture the real-world distribution of minority classes.

Second, inter-individual variability in Galvanic Skin Response presents a significant challenge. Factors such

as baseline skin conductance, hydration, and skin type can affect GSR readings. While Z-score normalization was applied to reduce inter-subject variance, this may not fully compensate for underlying physiological differences across users, potentially impacting generalizability.

Third, the primary dataset consisted of only 30 participants from a general population and involved four types of stress-inducing conditions (resting, academic, occupational, and social). While this provides a useful baseline for evaluating the system, the relatively small sample size and limited stress contexts may not fully represent real-world variability in stress triggers or demographic diversity.

Future studies should explore more robust sample sizes, diverse populations, and expanded physiological modalities (e.g., ECG, respiration) to improve personalization and reliability across contexts.

4. Conclusions

This study successfully developed and evaluated a real-time stress monitoring system that integrates Galvanic Skin Response (GSR) sensors with Internet of Things (IoT) technology and machine learning techniques, utilizing a Random Forest classifier enhanced by Synthetic Minority Over-sampling Technique (SMOTE) for class balancing. The system was trained and tested using a hybrid dataset combining primary GSR sensor readings from the general population and secondary data from the WESAD dataset.

The results demonstrate that the Random Forest model achieved a cross-validated accuracy of approximately 69%, with substantially improved performance across all stress categories, especially in detecting moderate and severe stress levels that are often underrepresented. Compared to manual threshold-based methods, the machine learning approach proved to be more flexible, adaptive, and capable of capturing subtle physiological patterns indicative of varying stress levels.

Descriptive statistical analysis confirmed that the system was effective across diverse demographic groups and different stress-inducing conditions. Validation against the Perceived Stress Scale (PSS) further supported the model's practical reliability, with a strong Cohen's Kappa agreement value of $\kappa = 0.82$, indicating substantial consistency between objective measurements and subjective perceptions of stress. Usability testing confirmed high acceptance, with users rating the system highly in terms of ease of use, simplicity, and interactivity.

Usability testing involving 30 participants confirmed that the system was accessible, easy to use, and interactive, demonstrating the potential for broader adoption beyond clinical settings. Moreover, the integration of real-time signal processing, user-friendly mobile and desktop interfaces, and predictive analytics marks a significant step forward from prior studies that relied solely on static datasets or lacked interactive deployment.

Nonetheless, limitations related to sample size, potential overfitting from synthetic data, and physiological variability highlight the need for further refinement. Future work should incorporate additional physiological modalities such as ECG and respiration, leverage deep learning models (e.g., CNN, LSTM) for improved temporal sensitivity, and conduct longitudinal studies in clinical or occupational contexts to assess chronic stress patterns. Integration with cloud-based storage and adaptive feedback systems could further enhance scalability, personalization, and real-time health monitoring capabilities.

Overall, this research offers a scalable, user-centered framework for objective stress detection and paves the way for more personalized and preventive mental health interventions in both individual and population-level applications.

Acknowledgements

This work was supported by a research grant from the Research and Community Service Body (LPPM IKTA) [Grant No. 98.1/IKTA/LPPM-KR/PL/2025].

References

[1] P. Chodavadia, I. Teo, D. Poremski, D. S. S. Fung, and E. A. Finkelstein, "Prevalence and economic burden of depression and anxiety symptoms among Singaporean adults: results from a 2022 web panel," *BMC Psychiatry*, vol. 23, p. 104, Feb. 2023, doi: https://doi.org/10.1186/s12888-023-04581-7.

[2] A. Jalali *et al.*, "Global prevalence of depression, anxiety, and stress in the elderly population: a systematic review and meta-analysis," *BMC Geriatr.*, vol. 24, p. 809, Oct. 2024, doi: https://doi.org/10.1186/s12877-024-05311-8.

[3] K. M. Harris, A. E. Gaffey, J. E. Schwartz, D. S. Krantz, and M. M. Burg, "The Perceived Stress Scale as a Measure of Stress: Decomposing Score Variance in Longitudinal Behavioral Medicine Studies," *Ann. Behav. Med. Publ. Soc. Behav. Med.*, vol. 57, no. 10, pp. 846–854, Apr. 2023, doi: https://doi.org/10.1093/abm/kaad015.

[4] F. B. Larsen, M. Lasgaard, M. V. Willert, and J. B. Sørensen, "Perceived stress across population segments characterized by differing stressor profiles—A latent class analysis," *PLOS ONE*, vol. 20, no. 1, p. e0316759, Jan. 2025, doi: https://doi.org/10.1371/journal.pone.0316759.

[5] A. Almadhor *et al.*, "Wrist-Based Electrodermal Activity Monitoring for Stress Detection Using Federated Learning," *Sensors*, vol. 23, no. 8, Art. no. 8, Jan. 2023, doi: 10.3390/s23083984.

[6] D. J. van der Mee, M. J. Gevonden, J. H. D. M. Westerink, and E. J. C. de Geus, "Validity of electrodermal activity-based measures of sympathetic nervous system activity from a wrist-worn device," *Int. J. Psychophysiol.*, vol. 168, pp. 52–64, Oct. 2021, doi: 10.1016/j.ijpsycho.2021.08.003.

[7] A. Georgas, A. Panagiotakopoulou, G. Bitsikas, K. Vlantoni, A. Ferraro, and E. Hristoforou, "Stress Monitoring in Pandemic Screening: Insights from GSR Sensor and Machine Learning Analysis," *Biosensors*, vol. 15, no. 1, Art. no. 1, Jan. 2025, doi: https://doi.org/10.3390/bios15010014.

[8] S. Andonov, S. Georgiev, G. Tsenov, and V. Mladenov, "Wearable Device for Pulse, Oximetry and Galvanic Skin Response Recordings for Neuromarketing," in *2024 13th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, Jun. 2024, pp. 01–06. doi: https://doi.org/10.1109/MOCAST61810.2024.10615512.

[9] A. Georgas, A. Panagiotakopoulou, G. Bitsikas, K. Vlantoni, A. Ferraro, and E. Hristoforou, "Stress Monitoring in Pandemic Screening: Insights from GSR Sensor and Machine Learning Analysis," *Biosensors*, vol. 15, no. 1, Art. no. 1, Jan. 2025, doi: https://doi.org/10.3390/bios15010014.

[10] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, in ICMI '18. New York, NY, USA: Association for Computing Machinery, Oct. 2018, pp. 400–408. doi: https://doi.org/10.1145/3242969.3242985.

[11] G. R. Reddy, J. J. Reddy, M. V. S. Reddy, I. Sravani, and V. K. A. Kumar, "Enhancing Stress Detection Employing Physiological Signals from the WESAD Dataset: A Machine Learning Approach with SMOTE," *Milest. Trans. Med. Technometrics*, vol. 2, no. 2, Art. no. 2, Dec. 2024, doi: https://doi.org/10.5281/zenodo.14566207.

[12] X. Yan, Z. Lin, Z. Lin, and B. Vucetic, "A Novel Exploitative and Explorative GWO-SVM Algorithm for Smart Emotion Recognition," *IEEE Internet Things J.*, vol. 10, no. 11, pp. 9999–10011, Jun. 2023, doi: https://doi.org/10.1109/JIOT.2023.3235356.

[13] D. S. Benita, A. S. Ebenezer, L. Susmitha, M. S. P. Subathra, and S. J. Priya, "Stress Detection Using CNN on the WESAD Dataset," in *2024 International Conference on Emerging Systems and Intelligent Computing (ESIC)*, Feb. 2024, pp. 308–313. doi: https://doi.org/10.1109/ESIC60604.2024.10481604.

[14] H.-S. Choi, "Emotion Recognition Using a Siamese Model and a Late Fusion-Based Multimodal Method in the WESAD Dataset with Hardware Accelerators," *Electronics*, vol. 14, no. 4, Art. no. 4, Jan. 2025, doi: https://doi.org/10.3390/electronics14040723.

[15] A. Garg *et al.*, "Inappropriate implantable cardioverter defibrillator discharge from sensing of external alternating current leak," *Journal Of Interventional Cardiac Electrophysiology*, vol. 7, no. 2. Kluwer Academic Publ, Van Godewijckstraat 30, 3311 GZ Dordrecht, Netherlands, pp. 181–184, Oct. 2002. doi: 10.1023/A:1020846007149.

[16] A. Mohammadi, M. Fakharzadeh, and B. Baraeinejad, "An Integrated Human Stress Detection Sensor Using Supervised Algorithms," *IEEE Sens. J.*, vol. 22, no. 8, pp. 8216–8223, Apr. 2022, doi: https://doi.org/10.1109/JSEN.2022.3157795.

[17] A. Al-Nafjan and M. Aldayel, "Anxiety Detection System Based on Galvanic Skin Response Signals," *Appl. Sci.*, vol. 14, no. 23, Art. no. 23, Jan. 2024, doi: https://doi.org/10.3390/app142310788.

[18] G. Mazumdar, K. Singh, and D. Sethia, "ML-based stress classification and detection using multimodal affect dataset," in *Advances in Networks, Intelligence and Computing*, CRC Press, 2024.

[19] E. Oliver and S. Dakshit, "Cross-Modality Investigation on WESAD Stress Classification," Feb. 26, 2025, *arXiv*: arXiv:2502.18733. doi: https://doi.org/10.48550/arXiv.2502.18733.

[20] M. A. Al Aleem, R. Mubarak, N. M. Salem, and I. Sadek, "A Deep Learning Approach Using WESAD Data for Multi-Class Classification with Wearable Sensors," in *2024 6th Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, Oct. 2024, pp. 194–197. doi: https://doi.org/10.1109/NILES63360.2024.10753228.

[21] G. Vos, K. Trinh, Z. Sarnyai, and M. Rahimi Azghadi, "Ensemble machine learning model trained on a new synthesized dataset generalizes well for stress prediction using wearable devices," *J. Biomed. Inform.*, vol. 148, p. 104556, Dec. 2023, doi: https://doi.org/10.1016/j.jbi.2023.104556.

[22] K. Dahal, B. Bogue-Jimenez, and A. Doblas, "Global Stress Detection Framework Combining a Reduced Set of HRV Features and Random Forest Model," *Sensors*, vol. 23, no. 11, Art. no. 11, Jan. 2023, doi: https://doi.org/10.3390/s23115220.

[23] Rawan S Abdulsadiq and Esther Rodriguez Villegas, "Frontiers | A comparative study in class imbalance mitigation when working with physiological signals," *Front. Digit. Health*, vol. 6, Mar. 2024, doi: https://doi.org/10.3389/fdgth.2024.1377165.

[24] G. R. Reddy, J. J. Reddy, M. V. S. Reddy, I. Sravani, and V. K. A. Kumar, "Enhancing Stress Detection Employing Physiological Signals from the WESAD Dataset: A Machine Learning Approach with SMOTE," *Milest. Trans. Med. Technometrics*, vol. 2, no. 2, Art. no. 2, Dec. 2024, doi: https://doi.org/10.5281/zenodo.14566207.

[25] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *J. Biomed. Inform.*, vol. 73, pp. 159–170, Sep. 2017, doi: https://doi.org/10.1016/j.jbi.2017.08.006.