HUMAN, EARTH AND FUTURE
Ital Publication

ISSN: 2785-2997

Available online at www.HEFJournal.org

Journal of
Human, Earth, and Future

Vol. 6, No. 3, September, 2025

Predicting Stress During Sleep from Biosensor Data: An Optimized Machine Learning Framework

Tran Anh Tuan [1, 2*], Dao Thi Thanh Loan [3]

[1] *School of Informatics, Walailak University, Nakhon Si Thammarat, 80160, Thailand.*

[2] *Informatics Innovation Center of Excellence (IICE), Walailak University, Nakhon Si Thammarat, 80160, Thailand.*

[3] *Department of Sciences, Dak Lak College of Pedagogy, Dak Lak, 63000, Vietnam.*

Received 09 June 2025; Revised 16 August 2025; Accepted 20 August 2025; Published 01 September 2025

Abstract

Objectives: This study aims to develop an optimized machine learning (ML) framework for predicting stress levels via physiological signals collected during sleep via Internet of Medical Things (IoMT)-enabled biosensors. The primary goal is to increase the accuracy and efficiency of stress prediction by identifying the most significant features that influence stress classification. *Method/Analysis*: The proposed framework employs two feature selection methods: particle swarm optimization combined with the whale optimization algorithm (PSO-WOA) and an enhanced version incorporating Lévy flight (PSO-WOA with Lévy flight). These methods are designed to reduce feature dimensionality while maximizing classification performance. A set of single (LR, KNN, NB, MLP, and SVM) and ensemble (RF, XGBoost, and Voting) classifiers are evaluated via 10-fold cross-validation. The Sleep-IoMT stress dataset, comprising biosensor-based physiological signals, was used for experimental validation. *Findings*: The framework achieved high classification accuracy across all the models, with all the classifiers exceeding 0.98 accuracy. Compared with the PSO-WOA, the PSO-WOA with the Lévy flight method demonstrated superior performance in terms of both feature selection quality and training time efficiency. The results confirm that effective feature selection significantly improves model accuracy and interpretability. *Novelty/Improvement*: This research introduces a hybrid approach for feature selection (PSO-WOA) in the context of stress prediction from sleep-related IoMT data. The integration of Lévy flight into the PSO-WOA enhances exploration capabilities and reduces premature convergence, offering a robust solution for real-world healthcare applications, e.g., mobile stress monitoring and early intervention systems.

Keywords: Stress Prediction; Feature Selection; PSO-WOA; PSO-WOA with Lévy Flight; Machine Learning; Single Classifiers; Ensemble Classifiers; Sleeping Monitor.

1. Introduction

Stress has emerged as a major contributor to mental and physical health problems worldwide [1]. Chronic stress is linked to various conditions (e.g., anxiety, depression, cardiovascular diseases, and weakened immunity) [2-4]. Among the many physiological effects of stress, its influence on sleep patterns is particularly profound. Sleep disturbances, e.g., fragmented sleep, rapid limb movement, irregular respiration, or heart rate variability, are often early indicators of psychological strain [5-7].

Recent advances in wearable sensors and the Internet of Medical Things (IoMT) have enabled continuous and nonintrusive monitoring of sleep-related physiological parameters. These smart medical devices provide valuable real-

* Corresponding author: tran.tu@wu.ac.th; tattinsphue@yahoo.com

http://dx.doi.org/10.28991/HEF-2025-06-03-010

time data streams, capturing key features such as the respiration rate, heart rate, eye movement, body temperature, and blood oxygen saturation [8, 9]. By leveraging these health-specific data, machine learning (ML) techniques can be employed to develop predictive models with clinical relevance, particularly in assessing and forecasting stress levels. A variety of ML methods have been explored for this purpose: Ciabattoni et al. [10] employed K-Nearest Neighbors (KNN), Nath et al. [11] used Random Forest (RF), Wu et al. [12] applied Logistic Regression (LR), and Rachakonda et al. [13] implemented a Deep Neural Network. Kumar et al. [14] used KNN with feature importance, while Anitha [15] developed a stacking ensemble model. Other approaches include Support Vector Machine (SVM) by Wahab et al. [16], Naive Bayes (NB) by Jayawickrama & Rupasingha [17], Multilayer Perceptron (MLP) by Rachakonda et al. [18], and RF by Shruthika & Rasheedha [19].

The integration of IoMT with ML represents a promising approach for personalized health monitoring and early intervention in stress-related conditions. However, raw sensor data are high-dimensional, noisy, and often redundant [20, 21]. Thus, feature selection plays a critical role in identifying the most informative attributes that contribute to stress classification. For instance, mutual information, ANOVA, and correlation-based approaches have been applied in stress detection tasks. Additionally, other methods such as nature-inspired metaheuristics, Boruta, and genetic algorithms have been explored in broader healthcare contexts, including HIV diagnosis, human resource prediction, depression detection using EEG data, biosensor signal analysis, and image-based grading tasks [22-28]. These traditional algorithms may struggle with local optima or overfitting when handling such complex datasets.

To address this, this study proposes a robust ML-based framework for stress prediction using sleep-related physiological data. The framework incorporates two feature selection methods: (1) a hybrid Particle Swarm Optimization-Whale Optimization Algorithm (PSO-WOA) and (2) its enhanced variant, the PSO-WOA with Lévy flight. These optimization techniques aim to identify the most informative subset of features that contribute to stress classification, thereby enhancing model accuracy and generalizability. The selected features are then used to train and test multiple classifiers, including both single (e.g., LR, KNN, NB, MLP, and SVM) and ensemble models (e.g., RF, Extreme Gradient Boosting: XGBoost, and a Voting Classifier). To ensure robustness and fairness in model evaluation, a 10-fold cross-validation strategy is employed during training, while key default hyperparameters are used for each classifier to maintain consistency and reproducibility. The performance of the proposed framework is assessed via standard evaluation metrics, including accuracy, precision, recall (sensitivity), F1-score, and specificity to capture different aspects of classification quality. Additionally, comparisons between the two feature selection approaches are conducted to determine their relative effectiveness in enhancing classifier performance. This study has the following contributions: 1) introduce an optimized machine-learning framework for stress prediction because of sleep-related physiological data collected through IoMT devices; and 2) propose intelligent optimization techniques for feature selection by introducing two hybrid metaheuristic algorithms.

The structure of this paper is as follows: Section 2 introduces the research methodology, including data collection, feature selection, and model evaluation. Section 3 presents experimental results, while Section 4 discusses key findings, limitations, and future research directions. Finally, Section 5 concludes the study with insights into the practical applications of the proposed approach.

2. Research Methodology

2.1. Data Collection

The dataset used in this study, referred to as the Sleep-IoMT stress dataset (630 samples and 8 columns), was obtained from Rachakonda et al. [18] and was derived from an IoMT-enabled device specifically designed to collect physiological signals associated with stress during sleep. This device captured a range of features considered relevant for stress prediction, including “Snoring Range” (SR), “Respiration Rate” (RR), “Body Temperature” (T), “Limb Movement rate” (LM), “Blood Oxygen level” (BO), “Eye movement during REM sleep” (REM), “number of hours of sleep” (SH), and “Heart Rate” (HR). The target variable, “Stress Level” (SL), was categorized into five classes: 0 - “Normal”, 1 - “Medium low”, 2 - “Medium”, 3 - “Medium high”, and 4 - “High”. These features were chosen for their clinical significance in assessing stress and serve as the foundation for the ML models developed in this research.

2.2. The Proposed Method

Figure 1 depicts the ML-based framework for stress prediction from the sleep-IoMT stress dataset, which includes physiological signals associated with stress during sleep. The framework has four significant phases, including *(1) data preprocessing and splitting*, *(2) feature selection for the training phase*, *(3) feature identification for testing phase*, and *(4) classification*. The input and output of the framework are the sleep-IoMT stress dataset and five stress levels (five classes), respectively.

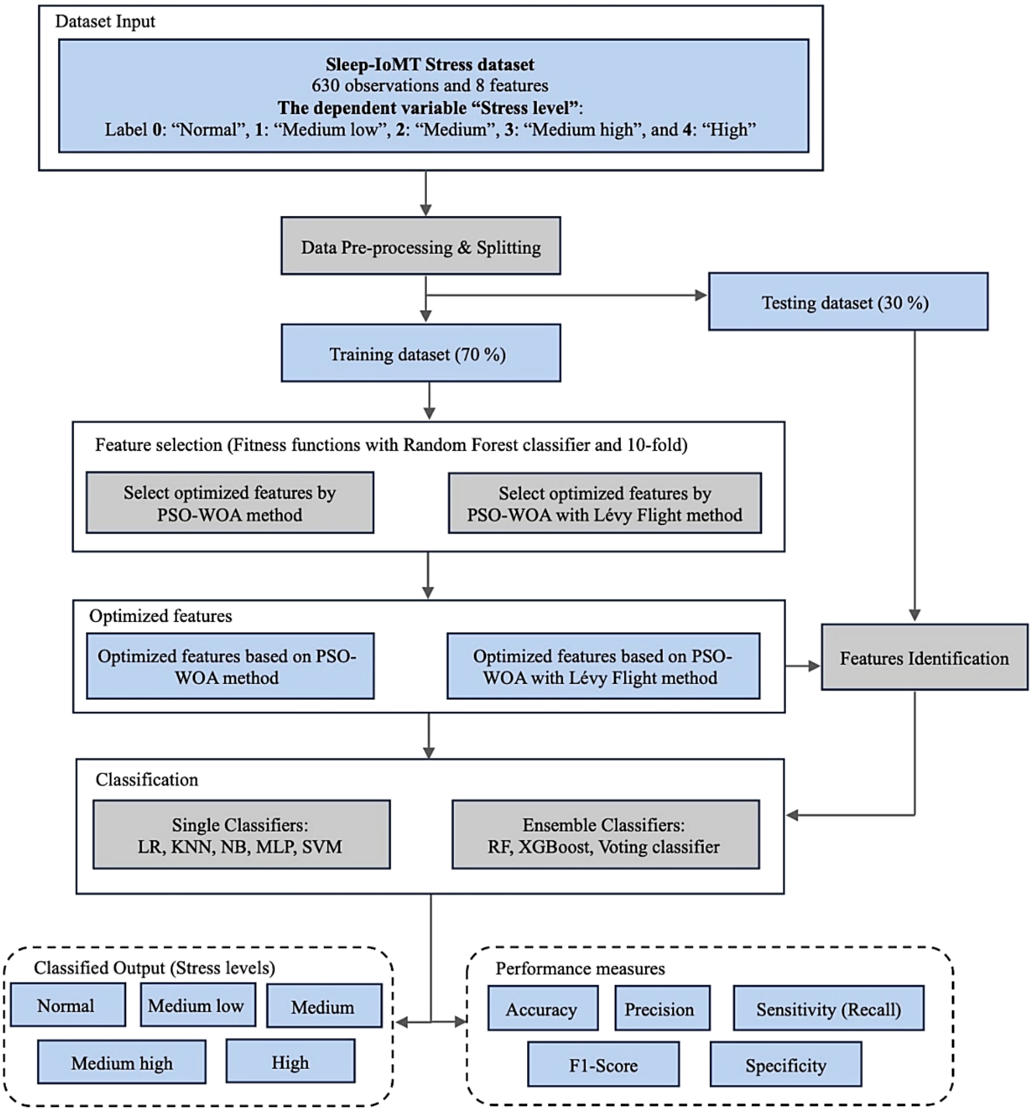


Figure 1. A framework based on ML for stress prediction from sleep-IoMT stress dataset

2.2.1. Data Pre-processing and Splitting

In the preprocessing stage, missing values within the dataset are addressed via mean imputation to ensure data completeness. To ensure uniformity in feature scaling, "MinMaxScaler" [29] is applied to normalize all feature values into the range of 0 - 1. Although the target values are numerical, label encoding was applied if categorical formats were present. The dataset is then divided into 70% for training and 30% for testing to allow for unbiased model evaluation.

2.2.2. Feature Selection for the Training Phase

To detect the most relevant physiological features for stress prediction, this study proposes two feature selection methods during the training phase: (1) particle swarm optimization combined with the whale optimization algorithm (PSO-WOA) method and (2) its enhanced variant, PSO-WOA with Lévy flight. The objective is to optimize feature subsets by maximizing classification performance while reducing feature dimensionality.

The PSO-WOA combines the social behavior of PSO, which mimics the movement of a swarm toward the best-known positions, with the exploitation strategy of the WOA, which simulates bubble-net hunting by humpback whales. This hybridization allows the algorithm to effectively balance global exploration and local exploitation. However, standard PSO-WOA may still suffer from premature convergence in complex search spaces. To address this, the Lévy flight was integrated into the PSO-WOA. Lévy flight employs stochastic long-distance jumps to escape local minima and explore a broader range of the search space [30-35]. For both methods, the fitness function is defined via the classification performance of a RF model evaluated via 10-fold cross-validation. Each candidate feature subset is represented as a binary vector, where 1 indicates the inclusion of a feature and 0 indicates exclusion. The subset of features with the highest average accuracy across 10-fold was selected for the subsequent classification phase. Processing steps to optimize feature subset by the PSO-WOA and PSO-WOA with Lévy flight methods are depicted in Figure 2.

```
Input:  Normalized training dataset with all features
Output: FP optimal feature subset obtained via the PSO-WOA method

        FL optimal feature subset obtained via the PSO-WOA with the Lévy flight method

1. Initialize population of candidate solutions (binary vectors representing feature subsets)

2. For each candidate in the population:
      a. Select features indicated by the binary vector
      b. The RF classifier is trained via 10-fold cross-validation
      c. Compute the classification accuracy as the fitness score

3. Update particle positions via PSO (velocity and position updates)

4. Apply the WOA encircling prey mechanism to refine the search

5. Lévy flight mutation is applied to introduce random long jumps (with a defined
   probability) to increase search diversity

6. Repeat steps until a predefined number of iterations or convergence for the following
   options

      a. Steps 2 - 4 for the PSO-WOA method

      b. Steps 2 - 5 for the PSO-WOA with Lévy flight method

7. Return the optimal feature subsets FP and FL with the highest fitness score
```

Figure 2. Optimized feature subsets by the PSO-WOA and the PSO-WOA with Lévy flight methods

Figure 2 illustrates the processing steps of the two optimization approaches: the PSO-WOA and the PSO-WOA with Lévy flight. In Step 1, a population of candidate solutions is initialized, where each individual is encoded as a binary vector representing a unique subset of features from the normalized training dataset. In Step 2 of Figure 2, for each candidate in the population, the selected features (indicated by binary values) are extracted. The RF classifier is then trained via 10-fold cross-validation, and the resulting classification accuracy is computed as the fitness score. Step 3 involves updating the particles' velocities and positions via the PSO algorithm, guiding the search toward promising regions in the feature space. In Step 4, the WOA is applied to refine the search by simulating the encircling prey behavior, which improves local exploitation. For the PSO-WOA with the Lévy flight approach, Step 5 introduces an additional mutation phase via Lévy flight with a predefined probability. This step introduces random long-distance jumps to increase the search diversity and mitigate the risk of premature convergence. These steps are repeated in Step 6 for a set number of iterations or until convergence is achieved. Specifically, Steps 2 to 4 are executed for the PSO-WOA method, whereas Steps 2 to 5 are followed for the PSO-WOA with the Lévy flight approach. Finally, in Step 7, the feature subsets that yield the highest classification accuracy, denoted as F_P (for the PSO-WOA) and F_L (for the PSO-WOA with Lévy flight), are selected as the optimal subsets.

2.2.3. Features Identification for the Testing Phase

In the testing phase, the feature subset identified during training is applied to the testing data to maintain consistency and avoid data leakage. Only the features selected by the PSO-WOA or the PSO-WOA with the Lévy flight methods are used for classification. The testing dataset is filtered to retain only these selected features, which were previously determined to have the most significant impact on model performance. The results of this phase are then used for stress prediction in the subsequent classification phase.

2.2.4. Classification Algorithms

In this study, two types of ML models are commonly used to predict stress from sleep data (single and ensemble ML classifiers). Single ML classifiers use a single, standalone model to perform classification or prediction tasks.

In contrast, ensemble ML classifiers combine the predictions of multiple models to improve overall performance. Single and ensemble ML classifiers have been deployed in various domains including the health domain [36]. In this study, we employ LR, KNN, NB, MLP, and SVM as single classifiers. For the ensemble classifiers, we use RF, XGBoost, and voting classifier. These classifiers are deployed with key hyperparameters by SK-learn in Python [29, 37, 38]. The key hyperparameters of these classifiers are depicted in Table 1.

Table 1. Key hyperparameters for single and ensemble classifiers

Type	Classifier	Key Hyperparameters
Single	LR	solver= 'lbfgs', C=1.0
	KNN	n_neighbors=5, metric= 'minkowski', weights= 'uniform'
	NB	GaussianNB with learned class priors
	MLP	hidden_layer_sizes=(100), activation= 'relu', solver= 'adam'
	SVM	kernel= 'rbf', C=1.0, gamma= 'scale'
Ensemble	RF	n_estimators=100, criterion= 'gini', bootstrap=True, max_features= 'sqrt'
	XGBoost	n_estimators=100, learning_rate=0.1, max_depth=3
	Voting classifier	Soft voting combining LR, RF, and XGBoost

2.2.5. Model Evaluation

To provide comprehensive model evaluation, popular measurement metrics (accuracy, precision, sensitivity (recall), F1-scores, and specificity) in various prediction problems [9, 39-41].

$$\text{Accuracy} = TP + TN/(TP + FP + TN + FN) \tag{1}$$

$$\text{Precision} = TP/(TP + FP) \tag{2}$$

$$\text{Sensitivity (Recall)} = TP/(TP + FN) \tag{3}$$

$$\text{F1} - \text{score} = (2 \times \text{P} \times \text{R})/(\text{P} + \text{R}) \tag{4}$$

$$\text{Specificity } = TN/(TN + FP) \tag{5}$$

where *TP* (true positive): the predicted value is positive and its positive. *FP* (false positive): the predicted value is positive and its negative. *TN* (true negative): the predicted value is negative and its negative. *FN* (false negative): the predicted value is negative but its positive.

3. Results

3.1. Characteristics of the Dataset

Among the 630 samples distributed across five stress levels, each level represented 126 samples (20.0%). The characteristics of the dataset at each stress level are summarized in Table 2.

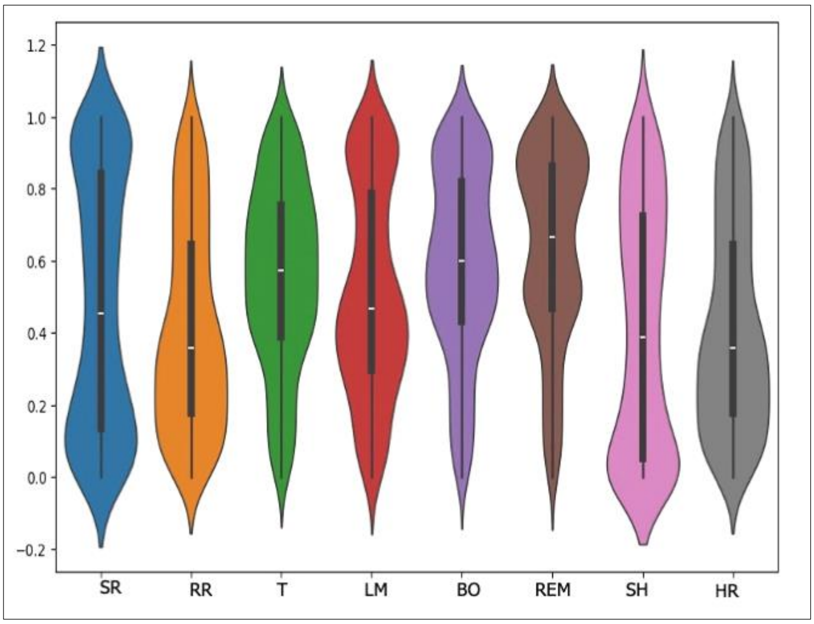
Table 2. Descriptive characteristics of the dataset. The breakdown of the total sample size is based on stress levels

Characteristics	All (n=630)	Stress levels					H-statistic
		Normal (n=126)	Medium low (n=126)	Medium (n=126)	Medium high (n=126)	High (n=126)	
SR							603.83***
median [IQR]	70.0 [52.5-91.25]	47.5 [46.25-48.75]	55.0 [52.5-57.5]	70.0 [65.0-75.0]	87.5 [83.75-91.25]	98.0 [97.0-99.0]	
RR							603.83***
median [IQR]	21.0 [18.5-25.0]	17.0 [16.5-17.5]	19.0 [18.5-19.5]	21.0 [20.5-21.5]	24.0 [23.0-25.0]	28.0 [27.0-29.0]	
T							603.83***
median [IQR]	93.0 [90.5-93.5]	97.5 [96.75-98.25]	95.0 [94.5-95.5]	93.0 [92.5-93.5]	91.0 [90.5-91.5]	87.5 [86.25-88.75]	
LM							603.83***
median [IQR]	11.0 [8.5-15.75]	6.0 [5.0-7.0]	9.0 [8.5-9.5]	11.0 [10.5-11.5]	14.5 [13.25-15.75]	18.0 [17.5-18.5]	
BO							603.83***
median [IQR]	91.0 [88.25-94.25]	96.0 [95.5-96.5]	93.5 [92.75-94.25]	91.0 [90.5-91.5]	89.0 [88.5-89.5]	85.0 [83.5-86.5]	
REM							603.83***
median [IQR]	90.0 [81.25-98.75]	70.0 [65.0-75.0]	82.5 [81.25-83.75]	90.0 [87.5-92.5]	97.5 [96.25-98.75]	102.5 [101.25-103.75]	
SH							608.34***
median [IQR]	3.5 [0.5-6.5]	8.0 [7.5-8.5]	6.0 [5.5-6.5]	3.5 [2.75-4.25]	1.0 [0.5-1.5]	0.0 [0.0-0.0]	
HR							603.83***
median [IQR]	62.5 [56.25-72.5]	52.5 [51.25-53.75]	57.5 [56.25-58.75]	62.5 [61.25-63.75]	70.0 [67.5-72.5]	80.0 [77.5-82.5]	

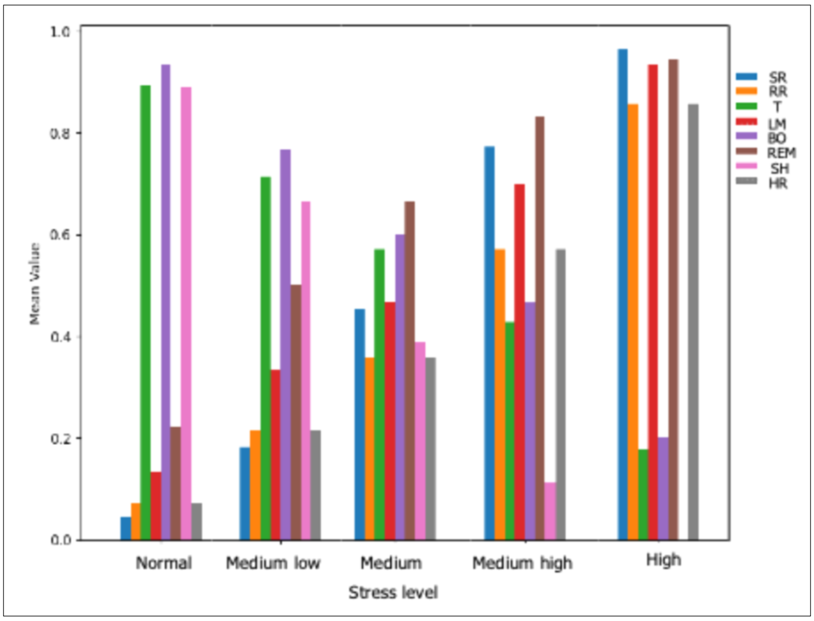
Note: *: $p < .005$; **: $p < .001$; ***: $p < .0001$

The results revealed clear physiological distinctions among the different stress levels. The SRs increased steadily from a median of 47.5 [IQR: 46.25–48.75] in the normal group to 98.0 [97.0–99.0] in the high-stress group. Similarly, the RR increased from 17.0 [16.5–17.5] in the normal group to 28.0 [27.0–29.0] in the high-stress group, and the HR increased from 52.5 [51.25–53.75] to 80.0 [77.5–82.5]. In contrast, T decreased from 97.5 [96.75–98.25] to 87.5 [86.25–88.75], and SH decreased sharply from 8.0 [7.5–8.5] in the normal group to 0.0 [0.0–0.0] in the high-stress group Additionally, results from the Kruskal–Wallis H-test (H-statistic) revealed statistically significant differences across stress levels for all variables, $H\ (4) \geq 603.83$, $p < 0.001$ all variables (Table 2).

To develop the stress prediction model, eight physiological features - SR, RR, T, LM, BO, REM, SH, and HR - were standardized. Each feature comprised 630 observations, with normalized values ranging from 0.0 to 1.0. Among these, REM exhibited the highest median importance (0.67 with IQR [0.47–0.86]), followed by BO (0.60 [0.43–0.82]) and T (0.57 [0.39–0.75]). In contrast, HR and RR had relatively lower median importance values (both at 0.36 [0.18–0.64]). The distributions also showed moderate dispersion, with standard deviations ranging from 0.252 for T to 0.352 for SR, suggesting variability in feature contributions across iterations or subjects (Figure 3-a). Among the eight features, SR, RR, LM, REM, and HR steadily increased from normal to high stress levels (normalized means from 0.05 to 0.96). In contrast, T, BO, and SH declined, indicating stress-related disruptions in homeostasis and rest (Figure 3-b).



(a) Normalized values across eight features



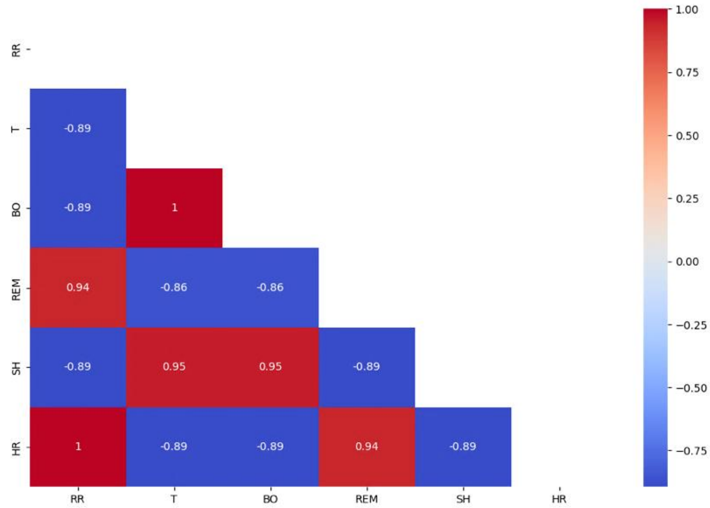
(b) Mean normalized values across five stress levels

Figure 3. Distribution of normalized values for features

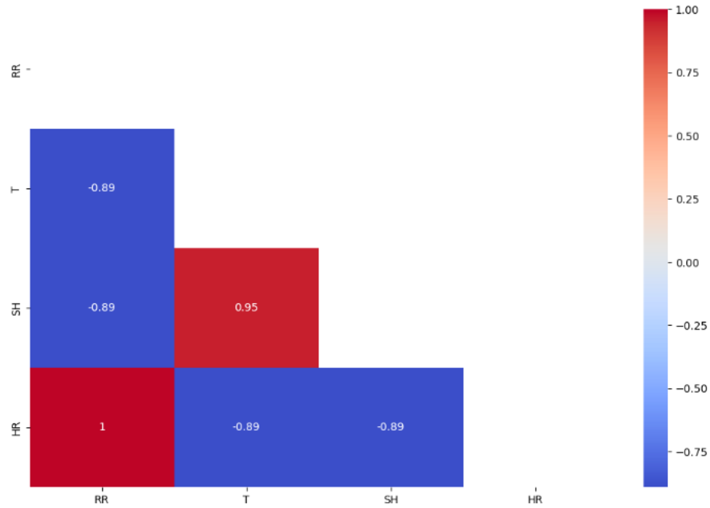
3.2. Feature Subset Optimization

In this study, feature subset optimization was performed via two methods: the PSO-WOA and the PSO-WOA with the Lévy flight. Each method aims to identify the most relevant physiological features from sleep-related sensor data for accurate stress level prediction. The resulting subsets differed in size and composition, reflecting the influence of the optimization strategy on feature selection. The PSO-WOA algorithm selects a total of six features: *RR*, *T*, *BO*, *REM*, *SH*, and *HR*. In contrast, the PSO-WOA algorithm augmented with Lévy flight produced a more compact subset of only four features: *RR*, *T*, *SH*, and *HR*. The consistent selection of RR, T, SH, and HR across both methods suggests that these features hold significant predictive value for stress classification.

For the PSO-WOA method, the correlation analysis among RR, T, BO, REM, SH, and HR reveals several strong relationships. RR and HR are perfectly correlated (r = 1.000), indicating potential redundancy. T and BO are also highly correlated (r = 0.998), whereas RR is strongly negatively correlated with both (r = -0.889). REM sleep positively correlates with RR and HR ($r \approx 0.936$), whereas SH sleep strongly positively correlates with T and BO ($r \approx$ 0.95), suggesting stable physiological conditions with longer sleep. Notably, REM and SH are strongly negatively correlated (r = -0.894), possibly reflecting disrupted sleep patterns under stress. These findings highlight interrelated physiological responses and suggest possible data overlap (Figure 4-a)



(a) PSO-WOA



(b) PSO-WOA with Lévy flight

Figure 4. Heap maps among optimized features by PSO-WOA and PSO-WOA with Lévy flight

For the PSO-WOA with the Lévy flight method, the correlation analysis among RR, T, SH, and HR reveals strong interrelationships. RR and HR are perfectly correlated (r = 1.000), indicating a tight physiological link or data redundancy. Both RR and HR are strongly negatively correlated with T and SH ($r \approx$ -0.889 to -0.892), suggesting that increased respiratory and heart rates are associated with lower body temperature and reduced sleep. In contrast, T and SH are strongly positively correlated (r = 0.955), indicating that higher temperatures align with longer sleep durations. These patterns suggest that elevated stress responses coincide with disrupted sleep and thermoregulation (Figure 4-b).

3.3. Model Performance

A comprehensive evaluation of model performance was conducted via multiple metrics, including accuracy, precision, recall, F1-score, specificity, and training time. The experimental results are described in Table 3.

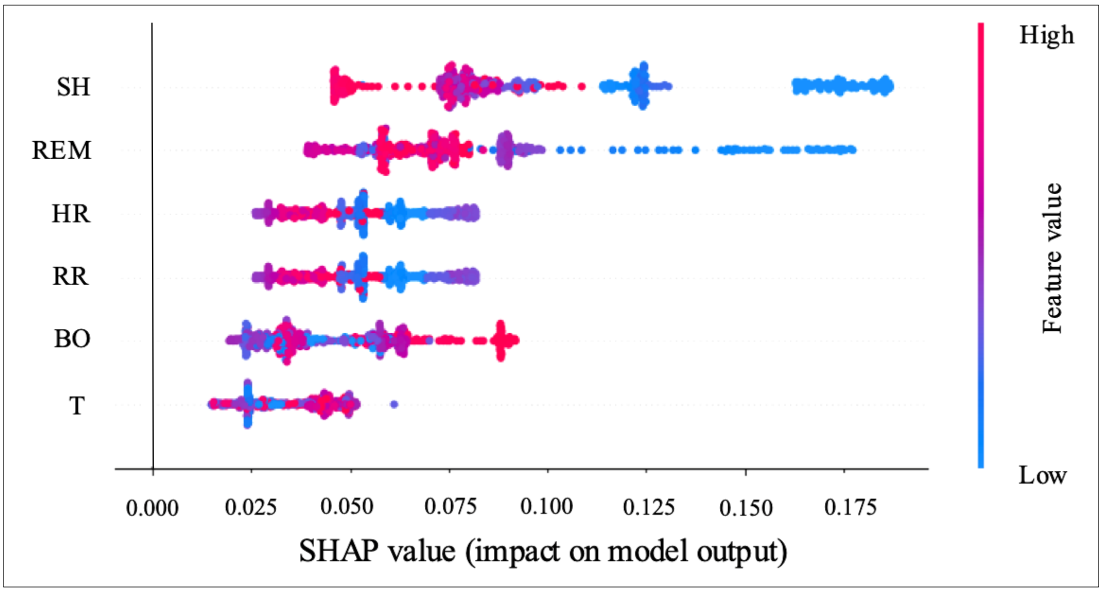
Table 3. A comparison of single and ensemble ML models based on the PSO-WOA and PSO-WOA with Lévy flight

Feature selection methods (Type of ML models)	ML model	Accuracy	Precision	Recall (Sensitivity)	F1-score	Specificity	Consuming time for training (s)
PSO-WOA -based (Single)	LR	1.000	1.000	1.000	1.000	1.000	0.01420
	KNN	1.000	1.000	1.000	1.000	1.000	0.02313
	NB	1.000	1.000	1.000	1.000	1.000	**0.00321**
	MLP	1.000	1.000	1.000	1.000	1.000	0.65171
	SVM	1.000	1.000	1.000	1.000	1.000	0.01714
PSO-WOA - based (Ensemble)	RF	0.995	0.995	0.994	0.995	0.999	0.20514
	XGBoost	0.984	0.984	0.984	0.984	0.996	0.16946
	Voting classifier	0.989	0.989	0.989	0.989	0.997	0.90654
PSO-WOA with Lévy flight-based (Single)	LR	1.000	1.000	1.000	1.000	1.000	0.01651
	KNN	1.000	1.000	1.000	1.000	1.000	0.01992
	NB	1.000	1.000	1.000	1.000	1.000	**0.00269**
	MLP	1.000	1.000	1.000	1.000	1.000	0.60563
	SVM	1.000	1.000	1.000	1.000	1.000	0.01448
PSO-WOA with Lévy flight-based (Ensemble)	RF	0.995	0.995	0.994	0.995	0.999	0.19425
	XGBoost	0.984	0.984	0.984	0.984	0.996	0.10946
	Voting classifier	0.989	0.989	0.989	0.989	0.997	0.50908

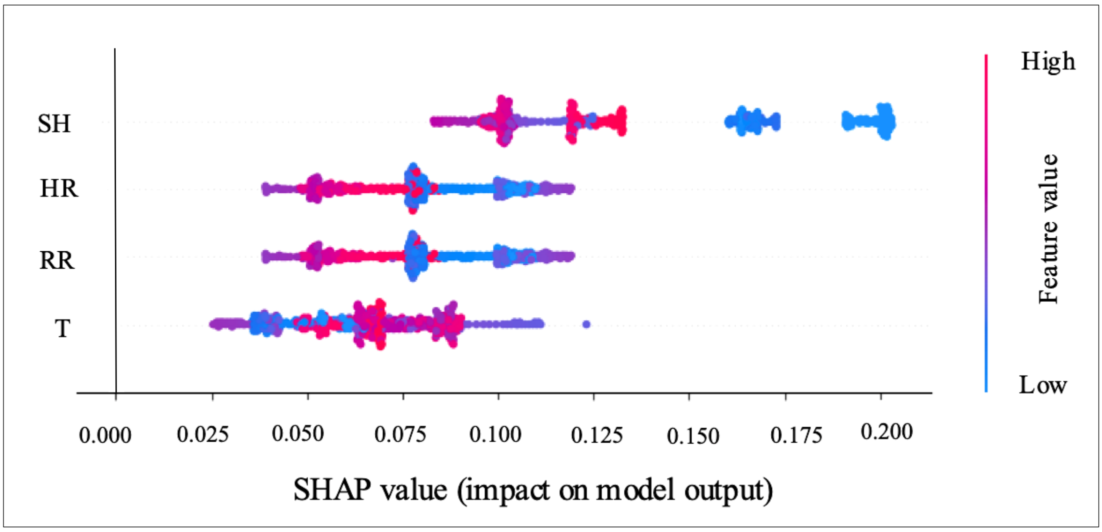
The experimental results demonstrated the exceptional performance of both the PSO-WOA and the PSO-WOA with Lévy flight optimization methods across a variety of ML classifiers. All single classifiers (LR, KNN, NB, MLP, and SVM) achieved perfect performance scores under both optimization approaches. These classifiers yielded accuracies, precisions, recalls, F1 score, and specificities of 1.000, indicating that the models were able to classify the stress levels of all the instances correctly. This level of performance signifies a perfect balance between false positives and false negatives and reflects the high generalizability of the optimized feature selection methods. In terms of computational efficiency, NB was the fastest classifier, requiring less than 0.003s for training under both the PSO-WOA and its Lévy-enhanced variant.

On the other hand, the MLP was the slowest among the single classifiers, with training times exceeding 0.6 seconds. Interestingly, the training time for the MLP and other models was slightly reduced under the PSO-WOA with the Lévy fligh variant, suggesting that the Lévy mechanism enhances the search process during optimization and reduces computational complexity. The Wilcoxon signed-rank test was employed to make comparisons between the two groups (the performance of PSO-WOA and PSO-WOA with Lévy flight feature selection methods) in terms of accuracy and runtime. From the test, no statistically significant difference in accuracy between the two approaches (p = 1.0), indicating comparable classification performance. However, a statistically significant difference in runtime was observed (p = 0.023), and the PSO-WOA with the Lévy flight method demonstrated improved computational efficiency. These results indicate that while both methods are equally effective in terms of prediction accuracy, the incorporation of Lévy flight offers a practical advantage by reducing execution time.

Across both SHAP plots Figure 5, SH consistently emerges as the most critical feature affecting model output. High SH values increase prediction in Figure 5a, while low SH values do so in Figure 5b, indicating a possible nonlinear relationship or different modeling contexts. HR and RR are also consistently impactful, with lower values contributing more to predictions in both plots. T shows minimal influence throughout. Overall, the model appears highly sensitive to variations in sleep and physiological signals, especially when they deviate from normal ranges.



(a) SHAP for PSO-WOA



(b) SHAP for PSO-WOA with Lévy flight

Figure 5. SHAP for PSO-WOA and PSO-WOA with Lévy flight

Ensemble models, including RF, XGBoost, and voting classifiers, also performed exceptionally well, achieving high accuracy (≥ 0.984) with slightly longer training durations cross two feature selection methods (the PSO-WOA and the PSO-WOA with Lévy flight). RF achieved the highest accuracy among ensembles at 0.995, with strong sensitivity (recall of 0.994) and near-perfect specificity (0.999). XGBoost achieved with 0.984 accuracy and excellent balanced metrics, whereas the voting classifier achieved an accuracy of 0.989. Notably, the ensemble models took longer to train than the single classifiers did, particularly the voting classifier, which required over 0.9 seconds under the PSO-WOA, though the training time decreased to approximately 0.5 seconds with the addition of Lévy flight (see Table 3 and Figure 6).

The results in Table 4 compare between our proposed method and other methods for stress prediction via various sensors, ML models, the number of extracted features, and the number of stress levels. Previous studies [10-12, 42, 43] have used fewer sensors (one to three sensors) such as GSR, respiration rate, body temperature, and HRV. These studies focused mainly on binary classification (two stress levels) and achieved accuracies ranging from 0.830 - 0.970. The models applied included KNN, SVM, and RF, demonstrating reasonable performance with a few extracted features. Subsequent studies by Jayawickrama & Rupasingha [17], and Shruthika & Rasheedha [19] expanded the sensor set to seven or eight physiological signals, covering limb movement, snoring range, body temperature, respiration rate, REM period, and others. Although the classification tasks became more complex (predicting up to five stress levels), these studies still maintained high accuracies, approximately 0.913-0.950.

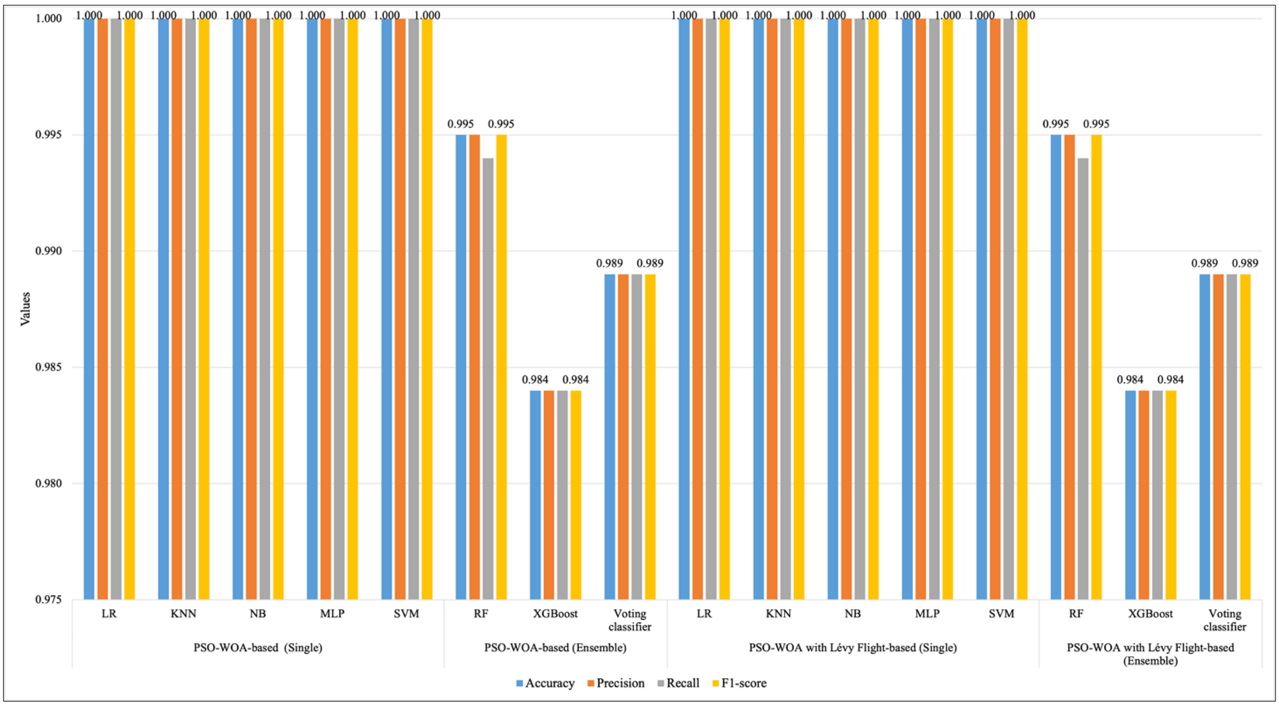


Figure 6. Comparison of performance among ML models based on feature selection methods

Table 4. A comparison of our proposed method with other methods

Author(s)	No. sensors	Sensors	ML model (best performance)	No. extracted features	No. Stress levels	Accuracy
Lawanot et al. [42]	2	Images, surveys	SVM	12	2	0.830
Ciabattonie et al. [10]	3	GSR, RR, BT	KNN	10	2	0.845
Jayawickrama et al. [17]	7	Limb movement, snoring range, body temperature, respiration rate, eye movement, blood oxygen level, heart rate	NB	7	2	0.913
Nath et al. [11]	2	GSR, PPG	RF	5	2	0.920
Shruthika & Rasheedha [19]	8	Heart rate, respiration rate, blood oxygen range, REM period, limb movement, body temperature, snoring range, sleep duration	RF	8	5	0.950
Rachakonda et al. [18] *	8	Heart rate, respiration rate, blood oxygen range, REM period, limb movement, body temperature, snoring range, sleep duration	MLP	8	5	0.960
Wu et al. [12]	1	HRV, salivary cortisol, Stress Response Inventory scores	LR	N/A	2	0.970
Rachakonda et al. [13]	3	Body temperature, steps taken, humidity	DNN	3	3	0.983
Wahab et al. [16] *	8	Heart rate, respiration rate, blood oxygen range, REM period, limb movement, body temperature, snoring range, sleep duration	SVM	8	5	0.994
Kumar et al. [14]*	8	Heart rate, respiration rate, blood oxygen range, REM period, limb movement, body temperature, snoring range, sleep duration	KNN	8	5	1.000
Anitha [15]*	8	Heart rate, respiration rate, blood oxygen range, REM period, limb movement, body temperature, snoring range, sleep duration	Stacking ensemble model	8	5	1.000
Our proposed method*	8	Heart rate, respiration rate, blood oxygen range, REM period, limb movement, body temperature, snoring range, sleep duration	NB	4	5	1.000

Note: Studies * conducted the same dataset.

The studies with an asterisk (*) in Table 4 used a consistent set of eight sleep-related physiological features (the same dataset) [14-16, 18]. All studies achieved accuracies ≥ 0.960. Among these five studies, our proposed method and the other two studies achieved accuracies of 1.000 with the NB, KNN, and the stacking ensemble models. However, our proposed method distinguished itself by reducing the number of extracted features from eight to four through feature selection. This demonstrates the effectiveness of optimal feature selection and lightweight models in achieving efficient and highly accurate stress prediction.

4. Discussion

This study proposed an optimized ML framework for predicting stress levels via sleep-related biosensor data. By integrating signals such as the respiratory rate (RR), heart rate (HR), REM sleep patterns, and body temperature with advanced feature selection and classification algorithms, the framework demonstrated a high degree of accuracy in predicting stress. The methodology highlights the importance of combining physiological signals with intelligent optimization techniques to increase the reliability and effectiveness of ML models in health-related applications.

Physiological features such as RR, HR, sleep hours (SH), and temperature (T) were strongly associated with stress levels. Notably, RR and HR were strongly correlated, suggesting a tight physiological coupling under stress. In addition, features such as REM sleep and blood oxygen (BO) also emerged as important contributors, reflecting the physiological disruptions that typically accompany elevated stress (see Figure 3 and Table 2). These findings are consistent with prior research linking autonomic nervous system activity, sleep quality, and metabolic changes to stress responses [40, 44, 45].

A key innovation in this research is the use of a hybrid PSO-WOA algorithm, which is further improved by incorporating the Lévy flight strategy. These enhancements significantly expand the search space diversity and mitigated the risk of becoming trapped in local optima. As a result, the proposed methods achieve superior feature selection by identifying the most relevant input attributes without sacrificing classification accuracy (≥ 0.996 all). This aligns with previous studies that mentioned the important role of feature selection methods [22, 46-48].

The proposed methods of feature selection revealed that SH, HR, and RR were consistently ranked as the most influential variables across models. This importance aligns with known biomarkers of psychological and physiological stress, indicating the model's ability to capture meaningful health patterns. This finding is consistent with recent studies that have linked stress factors [49-51].

In the context of evaluating model performance, this study emphasizes that relying solely on accuracy is insufficient, especially in multiclass biomedical prediction tasks (stress levels). Various metrics, e.g., sensitivity (true positive rate) and specificity (true negative rate), are particularly critical. High sensitivity ensures that individuals experiencing stress are correctly identified, which is vital for timely intervention and mental health support. On the other hand, high specificity helps reduce false alarms, avoiding unnecessary interventions or miscommunication. Additional metrics such as the F1-score further contribute to a well-rounded evaluation, offering insights into the model's balance between precision and recall across classes.

Despite these promising results, there are several limitations in this study. First, the dataset used in this study was balanced and was collected under standardized conditions, ensuring consistent measurements but potentially limiting generalizability to real-world environments. Second, individual factors such as age, sex, lifestyle, or medical history were not incorporated into the model, potentially affecting the interpretation of physiological responses.

5. Conclusion

This study proposed an optimized ML framework for predicting stress levels via physiological signals derived from sleep-related biosensor data. By leveraging advanced feature selection methods (the PSO-WOA and the PSO-WOA with the Lévy flight), the framework effectively identified the most influential features while maintaining high classification accuracy across various ML models (> 0.98 for all single and ensemble classifiers). The experimental results revealed the effective role of the feature selection method in enhancing model performance. The PSO-WOA with the Lévy flight method outperformed the PSO-WOA method in terms of training time efficiency. In addition, the proposed framework achieves better results than other methods do, validating its robustness and potential applicability. Key physiological indicators (SH, HR, and RR) highlighted by SHAP to confirm their clinical relevance in stress detection. These insights not only enhance model transparency but also offer valuable guidance for healthcare professionals in understanding and monitoring stress-related conditions. In the future, we aim to expand the dataset, incorporate additional behavioral parameters, and deploy the framework in real-world applications (e.g., mobile health monitoring systems) for stress management and early intervention.

6. Declarations

6.1. Author Contributions

Conceptualization, T.TA. and L.DTT.; methodology, T.TA. and L.DTT.; software, T.TA.; validation, T.TA.; formal analysis, T.TA.; resources, T.TA.; data curation, T.TA. and L.DTT.; writing—original draft preparation, T.TA.; writing—review and editing, T.TA. and L.DTT.; visualization, T.TA.; supervision, T.TA. and L.DTT. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available in the article.

6.3. Funding

This work was partially supported by Walailak University under the International Research Collaboration scheme (Contract Number WU-CIA-01801/2025).

6.4. Acknowledgments

The authors would like to thank the School of Informatics, IICE, CIA, Walailak University, for supporting this research.

6.5. Institutional Review Board Statement

Ethical review and approval were waived for this study because of *the use of secondary data* (Exemption with Approval Number: WUEC-24-322-01 by the Ethics Committee in Human Research Walailak University).

6.6. Informed Consent Statement

Not applicable.

6.7. Declaration of Competing Interest

The authors declare that there are no conflicts of interest concerning the publication of this manuscript. Furthermore, all ethical considerations, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7. References

[1] Wilson, D. (2025). Ranked: The World's Most and Least Stressed Countries, 2025. CEOWORLD magazine, New York, United States. Available online: https://ceoworld.biz/2025/01/13/ranked-the-worlds-most-and-least-stressed-countries-2025/ (accessed on August 2025).

[2] Marks, H., & King, L. M. (2020). Stress Symptoms. WedMD, New York, United States. Available online: https://www.webmd.com/balance/stress-management/stress-symptoms-effects_of-stress-on-the-body#1 (accessed on August 2025).

[3] Tuan, T. A., Trang, L. T. T., An, T. D., Nghia, N. H., & Loan, D. T. T. (2024). An Analysis of Machine Learning for Detecting Depression, Anxiety, and Stress of Recovered COVID-19 Patients. Journal of Human, Earth, and Future, 5(1), 1–18. doi:10.28991/HEF-2024-05-01-01.

[4] Tuan, T. A., Nghia, N. H., An, T. D., & Loan, D. T. T. (2024). Exploring Mental Stress Expressions in Online Communities: A Subreddit Analysis. Journal of Human, Earth, and Future, 5(2), 131–150. doi:10.28991/HEF-2024-05-02-01.

[5] Howarth, N. E., & Miller, M. A. (2024). Sleep, Sleep Disorders, and Mental Health: A Narrative Review. Heart and Mind, 8(3), 146–158. doi:10.4103/hm.hm-d-24-00030.

[6] Pekgor, M., Algin, A., & Toros, T. (2025). RFID-embedded mattress for sleep disorder detection for athletes in sports psychology. Scientific Reports, 15(1), 14697. doi:10.1038/s41598-025-96311-0.

[7] Nilsson, P. M., Nilsson, J. Å., Hedblad, B., & Berglund, G. (2001). Sleep disturbance in association with elevated pulse rate for prediction of mortality - Consequences of mental strain? Journal of Internal Medicine, 250(6), 521–529. doi:10.1046/j.1365-2796.2001.00913.x.

[8] Arumugam, G. R. M., Anbananthen, K. S. M., & Muthaiyah, S. (2025). Improving Patient Turnaround Time in Malaysian Hospitals with Real-Time IoT BLE Location Tracking. Journal of Human, Earth, and Future, 6(1), 115–130. doi:10.28991/HEF-2025-06-01-08.

[9] Kajornkasirat, S., Sawangwong, C., Puangsuwan, K., Chanapai, N., Phutthamongkhon, W., & Puttinaovarat, S. (2025). Integrating AI-Driven Predictive Analytics in Wearable IoT for Real-Time Health Monitoring in Smart Healthcare Systems. Applied Sciences (Switzerland), 15(8), 4400. doi:10.3390/app15084400.

[10] Ciabattoni, L., Ferracuti, F., Longhi, S., Pepa, L., Romeo, L., & Verdini, F. (2017). Real-time mental stress detection based on smartwatch. 2017 IEEE International Conference on Consumer Electronics, ICCE 2017, 110–111. doi:10.1109/ICCE.2017.7889247.

[11] Nath, R. K., Thapliyal, H., & Caban-Holt, A. (2020). Validating Physiological Stress Detection Model Using Cortisol as Stress Bio Marker. 2020 IEEE International Conference on Consumer Electronics (ICCE), 1–5. doi:10.1109/icce46568.2020.9042972.

[12] Wu, W., Pirbhulal, S., Zhang, H., & Mukhopadhyay, S. C. (2019). Quantitative Assessment for Self-Tracking of Acute Stress Based on Triangulation Principle in a Wearable Sensor System. IEEE Journal of Biomedical and Health Informatics, 23(2), 703–713. doi:10.1109/JBHI.2018.2832069.

[13] Rachakonda, L., Mohanty, S. P., Kougianos, E., & Sundaravadivel, P. (2019). Stress-Lysis: A DNN-Integrated Edge Device for Stress Level Detection in the IoMT. IEEE Transactions on Consumer Electronics, 65(4), 474–483. doi:10.1109/TCE.2019.2940472.

[14] Kumar, A., Sharma, N., Chauhan, R., Joshi, K., Mitra, D., & Madaan, V. (2024). Predicting Stress Levels in Sleep Patterns using Machine Learning Models and a Smart Yoga Pillow. 2024 7th International Conference on Contemporary Computing and Informatics (IC3I), 516–522. doi:10.1109/ic3i61595.2024.10828782.

[15] Anitha, K. L. (2025). Stacking Ensemble Model for Stress Detection using Sleeping Habits. 2025 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), 1–4. doi:10.1109/iitcee64140.2025.10915516.

[16] Wahab, O., Adda, M., & Zrira, N. (2024). SereniSens: a Multimodal AI Framework with LLMs for Stress Prediction through Sleep Biometrics. Procedia Computer Science, 251, 342–349. doi:10.1016/j.procs.2024.11.119.

[17] Jayawickrama, J. G., & Rupasingha, R. A. H. M. (2023). Human Stress Detection Based on Sleeping Habits Using Machine Learning Algorithms. 2023 3rd International Conference on Advanced Research in Computing (ICARC), 7–12. doi:10.1109/icarc57651.2023.10145663.

[18] Rachakonda, L., Bapatla, A. K., Mohanty, S. P., & Kougianos, E. (2021). SaYoPillow: Blockchain-Integrated Privacy-Assured IoMT Framework for Stress Management Considering Sleeping Habits. IEEE Transactions on Consumer Electronics, 67(1), 20–29. doi:10.1109/TCE.2020.3043683.

[19] Shruthika, L., & Rasheedha, A. (2025). Machine learning strategies for human stress level prediction: a comparative performance analysis. IET Conference Proceedings, 2024(23), 235–239. doi:10.1049/icp.2024.4429 9.

[20] Brabin, D. R., Agitha, W., Lakshmi, P. S., & Kumar, K. K. (2025). Advanced Deep Learning for IoT Sensor Data Processing in Autonomous Vehicle Navigation Systems. AI's Role in Enhanced Automotive Safety. IGI Global Scientific Publishing, Hershey, United States. doi:10.4018/979-8-3373-0442-7.ch001.

[21] Krishnamurthi, R., Kumar, A., Gopinathan, D., Nayyar, A., & Qureshi, B. (2020). An overview of IoT sensor data processing, fusion, and analysis techniques. Sensors (Switzerland), 20(21), 1–23. doi:10.3390/s20216076.

[22] Benghazouani, S., Nouh, S., & Zakrani, A. (2025). Optimizing breast cancer diagnosis: Harnessing the power of nature-inspired metaheuristics for feature selection with soft voting classifiers. International Journal of Cognitive Computing in Engineering, 6, 1–20. doi:10.1016/j.ijcce.2024.09.005.

[23] Manikandan, G., Pragadeesh, B., Manojkumar, V., Karthikeyan, A. L., Manikandan, R., & Gandomi, A. H. (2024). Classification models combined with Boruta feature selection for heart disease prediction. Informatics in Medicine Unlocked, 44, 101442. doi:10.1016/j.imu.2023.101442.

[24] Saha, R., Malviya, L., Jadhav, A., & Dangi, R. (2024). Early stage HIV diagnosis using optimized ensemble learning technique. Biomedical Signal Processing and Control, 89, 105787. doi:10.1016/j.bspc.2023.105787.

[25] Qi, Z., Yang, S., He, X., & Shi, D. (2024). Research on Human Resource Management and Prediction Model Based on Neural Network Optimized by Genetic Algorithm. 2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA), 38, 2049–2053. doi:10.1109/icipca61593.2024.10709104.

[26] Hassan, M., & Kaabouch, N. (2024). Impact of Feature Selection Techniques on the Performance of Machine Learning Models for Depression Detection Using EEG Data. Applied Sciences (Switzerland), 14(22), 10532. doi:10.3390/app142210532.

[27] Challhua, R., Prati, R., & Champi, A. (2024). Feature engineering and machine learning for electrochemical detection of rabies virus in graphene-based biosensors. Microchemical Journal, 204, 111074. doi:10.1016/j.microc.2024.111074.

[28] Habib, S., Khan, I., Aladhadh, S., Islam, M., & Khan, S. (2022). External Features-Based Approach to Date Grading and Analysis with Image Processing. Emerging Science Journal, 6(4), 694–704. doi:10.28991/ESJ-2022-06-04-03.

[29] Scikit-learn Developers. (2025). 3.1. Cross-validation: evaluating estimator performance. Available online: https://scikit-learn.org/stable/modules/cross_validation.html (accessed on August 2025).

[30] Jamil, M., & Zepernick, H.-J. (2013). Lévy Flights and Global Optimization. Swarm Intelligence and Bio-Inspired Computation, 49–72. Elsevier, Amsterdam, Netherlands. doi:10.1016/b978-0-12-405163-8.00003-x.

[31] Viswanathan, G. M., Afanasyev, V., Buldyrev, S. V., Havlin, S., da Luz, M. G. E., Raposo, E. P., & Stanley, H. E. (2000). Lévy flights in random searches. Physica A: Statistical Mechanics and Its Applications, 282(1–2), 1–12. doi:10.1016/s0378-4371(00)00071-6.

[32] Reynolds, A. M., & Rhodes, C. J. (2009). The Lévy flight paradigm: Random search patterns and mechanisms. Ecology, 90(4), 877–887. doi:10.1890/08-0153.1.

[33] Marini, F., & Walczak, B. (2015). Particle swarm optimization (PSO). A tutorial. Chemometrics and Intelligent Laboratory Systems, 149, 153–165. doi:10.1016/j.chemolab.2015.08.020.

[34] Mohammed, H. M., Umar, S. U., & Rashid, T. A. (2019). A Systematic and Meta-Analysis Survey of Whale Optimization Algorithm. Computational Intelligence and Neuroscience, 2019(1), 8718571. doi:10.1155/2019/8718571.

[35] Nasiri, J., & Khiyabani, F. M. (2018). A whale optimization algorithm (WOA) approach for clustering. Cogent Mathematics & Statistics, 5(1), 1483565. doi:10.1080/25742558.2018.1483565.

[36] Chung, J., & Teo, J. (2023). Single classifier vs. ensemble machine learning approaches for mental health prediction. Brain Informatics, 10(1), 1–10. doi:10.1186/s40708-022-00180-6.

[37] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research, 12, 2825-2830.

[38] Bates, S., Hastie, T., & Tibshirani, R. (2024). Cross-validation: what does it estimate and how well does it do it?. Journal of the American Statistical Association, 119(546), 1434-1445. doi:10.1080/01621459.2023.2197686.

[39] Nuankaew, P., Sorat, J., Intajak, J., Inta, J., & Nuankaew, W. S. (2025). AI for Healthcare: A Classification Model for Personalized Premenstrual Symptoms and Depressive Crisis Risk Tracking Using Data Analytics and Machine Learning. 2025 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering, 299–304. doi:10.1109/ectidamtncon64748.2025.10962042.

[40] Monga, S. V., Chaudhary, R. K., Kumar, D., Jahanavi, & Jain, V. (2024). Predicting Mental Health Stress Levels Through Sleep Patterns using Ensemble Models and Explainable AI. 2024 3rd Edition of IEEE Delhi Section Flagship Conference (DELCON), 1–7. doi:10.1109/delcon64804.2024.10866636.

[41] Yang, X., Yang, Z. X., Ye, J. Y., & Chen, Y. Y. (2025). Kernel-free quadratic proximal support vector machine with Lp-norm regularization. Engineering Applications of Artificial Intelligence, 151, 110658. doi:10.1016/j.engappai.2025.110658.

[42] Lawanot, W., Inoue, M., Yokemura, T., Mongkolnam, P., & Nukoolkit, C. (2019). Daily Stress and Mood Recognition System Using Deep Learning and Fuzzy Clustering for Promoting Better Well-Being. 2019 IEEE International Conference on Consumer Electronics, ICCE 2019, 1–6. doi:10.1109/ICCE.2019.8661932.

[43] Haque, Y., Zawad, R. S., Rony, C. S. A., Al Banna, H., Ghosh, T., Kaiser, M. S., & Mahmud, M. (2024). State-of-the-art of stress prediction from heart rate variability using artificial intelligence. Cognitive Computation, 16(2), 455-481. doi:10.1007/s12559-023-10200-0.

[44] Hossain, A. A., Ahamed, I. U., Gupta, U. Das, Anika, A. N., Islam, A., & Saha, U. (2024). Exploring Sleep Patterns and Behavioral Components for Stress Detection: A Machine Learning Approach. 2024 International Symposium on Networks, Computers and Communications, ISNCC 2024, 1–6. doi:10.1109/ISNCC62547.2024.10758990.

[45] Dasgupta, A., & Basu, A. (2025). Measurement and Instrumentation for Mental Stress Detection in Humans. IEEE Instrumentation and Measurement Magazine, 28(2), 10–17. doi:10.1109/MIM.2025.10946259.

[46] Sadeghian, Z., Akbari, E., Nematzadeh, H., & Motameni, H. (2025). A review of feature selection methods based on meta-heuristic algorithms. Journal of Experimental & Theoretical Artificial Intelligence, 37(1), 1–51. doi:10.1080/0952813X.2023.2183267.

[47] Alyasiri, O. M., Cheah, Y.-N., Zhang, H., Al-Janabi, O. M., & Abasi, A. K. (2024). Text classification based on optimization feature selection methods: a review and future directions. Multimedia Tools and Applications, 84(15), 14187–14233. doi:10.1007/s11042-024-19769-6.

[48] Wang, H., Liang, Q., Hancock, J. T., & Khoshgoftaar, T. M. (2024). Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. Journal of Big Data, 11(1), 44. doi:10.1186/s40537-024-00905-w.

[49] Ju, E., Guo, Y., Park, J. I., Kim, J., Qu, A., & Lee, J. A. (2025). Sleep Quality of Persons with Dementia and Family Caregivers in Korean Americans: Wearable Technology to Study the Dyadic Association. Western Journal of Nursing Research, 47(5), 404–412. doi:10.1177/01939459251323675.

[50] Sweeten, B. L. W., Sutton, A. M., Wellman, L. L., & Sanford, L. D. (2020). Predicting stress resilience and vulnerability: Brain-derived neurotrophic factor and rapid eye movement sleep as potential biomarkers of individual stress responses. Sleep, 43(1), 199. doi:10.1093/sleep/zsz199.

[51] Mateo-Reyes, R., Cruz-Albarran, I. A., & Morales-Hernandez, L. A. (2025). Intelligent Stress Detection Using ECG Signals: Power Spectrum Imaging with Continuous Wavelet Transform and CNN. Journal of Experimental and Theoretical Analyses, 3(1), 6. doi:10.3390/jeta3010006.