MethodsX 14 (2025) 103205

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/methodsx

From lab to real-life: A three-stage validation of wearable technology for stress monitoring ☆

Basil A. Darwish [a,b,*], Shafiq Ul Rehman [c], Ibrahim Sadek [a], Nancy M. Salem [a], Ghada Kareem [d], Lamees N. Mahmoud [a]

[a] Biomedical Engineering Department, Faculty of Engineering, Helwan University, Egypt
[b] Computer Science, Artificial Intelligence Programme, University of Hertfordshire hosted by Global Academic Foundation, Egypt
[c] College of Information Technology, Kingdom University, Kingdom of Bahrain
[d] Department of Biomedical Engineering, Higher Technological Institute, 10th Ramadan City, Egypt

ARTICLE INFO

Method name:
Lab-to-Real Multimodal Stress Detection Framework

Keywords:
Electrocardiogram (ECG)
Electrodermal activity (EDA)
Ensemble methods
Psychological stress detection
Respiration (RESP)
Wearable devices

ABSTRACT

Stress negatively impacts health, contributing to hypertension, cardiovascular diseases, and immune dysfunction. While conventional diagnostic methods, such as self-reported questionnaires and basic physiological measurements, often lack the objectivity and precision needed for effective stress management, wearable devices present a promising avenue for early stress detection and management. This study conducts a three-stage validation of wearable technology for stress monitoring, transitioning from controlled experimental data to real-life scenarios. Using the controlled WESAD dataset, binary and five-class classification models were developed, achieving maximum accuracies of 99.78 %±0.15 % and 99.61 %±0.32 %, respectively. Electrocardiogram (ECG), Electrodermal Activity (EDA), and Respiration (RESP) were identified as reliable stress biomarkers. Validation was extended to the SWEET dataset, representing real-life data, to confirm generalizability and practical applicability. Furthermore, commercially available wearables supporting these modalities were reviewed, providing recommendations for optimal configurations in dynamic, real-world conditions. These findings demonstrate the potential of multimodal wearable devices to bridge the gap between controlled studies and practical applications, advancing early stress detection systems and personalized stress management strategies.

- Stress detection methods were validated using multimodal wearable data in controlled (WESAD) and real-life (SWEET) datasets.
- Commercial wearable technologies were reviewed, offering insights into their applicability for practical stress monitoring.

Specifications table

Subject area:	Engineering
More specific subject area:	Healthcare
Name of your method:	Lab-to-Real Multimodal Stress Detection Framework
Name and reference of original method:	NA

(*continued on next page*)

☆ **Related research article:** None.
* Corresponding author.
E-mail address: Basil.Ahmed@h-eng.helwan.edu.eg (B.A. Darwish).

https://doi.org/10.1016/j.mex.2025.103205
Received 19 December 2024; Accepted 4 February 2025
Available online 5 February 2025

Resource availability:	The WESAD dataset, utilized in this study, is publicly accessible and can be downloaded from the following link: https://ubicomp.eti.uni-siegen.de/home/datasets/icmi18/. The SWEET dataset was obtained under a formal agreement with IMEC, OnePlanet Research Center, Netherlands. Requests for access to the SWEET dataset should be directed to Prof. Chris Van Hoof at IMEC, OnePlanet Research Center, Netherlands.

Background

Stress, a condition of mental strain or pressure due to upsetting conditions, is a major contributor to human physiology and pathophysiology. It has been linked to several conditions, such as autoimmune diseases, metabolic syndrome, sleep disorders, and suicidal thoughts and inclination [1]. Chronic stress affects over 70 % of Americans and has far-reaching consequences on physical, mental, and social health. It is associated with the development of cancer, cardiovascular diseases, depression, and diabetes, underscoring its detrimental effects on overall well-being [2–4].

The brain's response to stress involves a complex neural network, requiring precise neuroanatomical processing to identify and interpret threats [1]. Traditional diagnostic tools, such as self-reported questionnaires and basic physiological measures, lack the objectivity and accuracy needed for effective stress management [5]. Additionally, these methods are limited in their capacity for continuous, long-term monitoring, a critical requirement for addressing chronic stress and its impacts.

Emerging evidence highlights the role of stress in cardiovascular disease, driven by mechanisms like increased sympathetic activation, elevated blood pressure, and inflammatory responses. These changes are particularly significant in individuals with pre-existing conditions, emphasizing the need for robust, real-time stress detection methods [6].

Artificial Intelligence (AI) and Machine Learning (ML) offer promising avenues for stress detection, achieving high accuracy in distinguishing normal versus abnormal brain states, including those observed in post-traumatic stress disorder (PTSD) [1,7]. Recent advancements in ML techniques have further enhanced their predictive capabilities, making them valuable tools for stress detection [8].

Despite these advancements, several challenges and limitations exist in implementing ML for stress detection. Issues such as data privacy, the requirement for large and diverse datasets, and potential biases in algorithms need to be addressed [9]. Given that physiological signals have been demonstrated to be dependable indicators of stress [10], integrating multi-modal data (e.g., physiological signals, behavioral data, and environmental factors) could improve the robustness and accuracy of ML models [11].

The potential benefits of accurate stress detection using ML are extensive, including early intervention, personalized treatment plans, and improved mental health outcomes. To fully realize these benefits, future research should focus on refining ML algorithms, addressing ethical concerns, and developing user-friendly applications for both clinicians and patients [12]. In summary, while the application of ML in stress detection is promising, continuous research and development are crucial to overcome existing limitations and fully leverage these technologies to improve mental health care [13].

Physiological signals, including heart rate, electrodermal activity (EDA), and respiratory patterns, have been validated as reliable stress biomarkers [14–21]. Integrating these multimodal data streams with ML models offers the potential for early intervention, personalized treatment plans, and improved mental health outcomes. Nonetheless, addressing technical, ethical, and usability challenges is essential to fully leverage these technologies in real-world applications.

Given the critical importance of accurate stress detection, this study focuses on a three-stage validation of wearable technology for stress monitoring, using the WESAD dataset for controlled experimental conditions and the SWEET dataset to represent real-life scenarios. By integrating multimodal physiological signals and leveraging ML techniques, this methodology aims to bridge the gap between laboratory research and practical applications, advancing the field of stress detection and management.

Method details

To address the research gaps identified earlier, this study aims to explore the feasibility of robust psychological stress detection employing systematic signal segmentation and feature extraction to comprehensively characterize physiological responses, a rigorous machine learning pipeline involving feature selection, diverse classifiers, and hyperparameter optimization, a meticulous performance evaluation utilizing K-fold cross-validation (K-CV) as well as binary and multi-class detection. A comprehensive overview of our workflow is presented in Fig. 1.

Data acquisition

This study employs the Wearable Stress and Affect Detection (WESAD) dataset, a publicly available resource for stress classification research (https://ubicomp.eti.uni-siegen.de/home/datasets/icmi18/) [15], for the purposes of training, hyperparameter optimization, and testing. Subsequently, the stress in the work environment (SWEET) [10] dataset is employed for an additional round of training and testing with prior permission as a proof of concept, using the optimized hyperparameters derived from the WESAD dataset. Within WESAD, data for 15 subjects was collected across three experimental conditions designed to elicit varying stress levels: baseline, stress, and amusement. This study utilizes an electrocardiogram (ECG), electrodermal activity (EDA), and respiration (RESP) collected from the chest-worn RespiBAN Professional device, all sampled at 700 Hz. The labels used in this study were from the Positive and Negative Affect Schedule questionnaire (PANAS) available in WESAD, more specifically the 21st item (Stressed) with its five possible responses (1 = Not at all, 2 = *A* little bit, 3 = Somewhat, 4 = Very much, 5 = Extremely) in case of multi-class. For

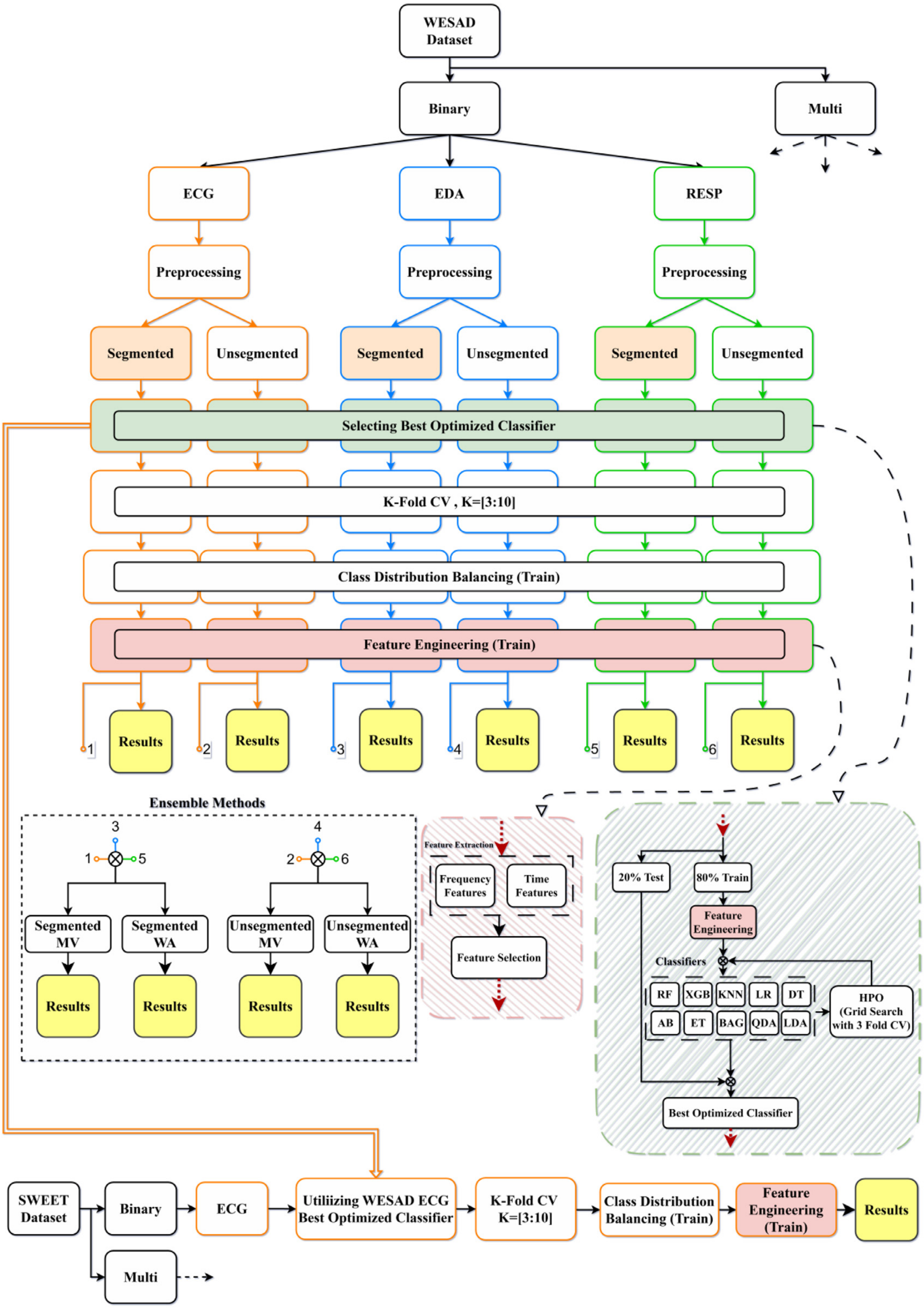

Fig. 1. Comprehensive binary and multi-class stress classification workflow for preprocessing, segmentation, classifier selection with hyperparameter optimization and feature engineering of ECG, EDA, and RESP modalities from the WESAD dataset and further verification using the SWEET dataset using WESAD hyperparameters.

binary classification, responses of ’Not at all’ were considered class 0 (no stress) with remaining response levels considered class 1 (stress).

WESAD subject information

- Demographics: The dataset includes data from 15 participants (12 males, 3 females) with a mean age of 27.5 ± 2.4 years.
- Inclusion criteria: Graduate students at their research facility.
- Exclusion criteria: Pregnancy, heavy smoking, mental disorders, chronic and cardiovascular diseases.

WESAD experimental procedures

- Study Design: The study was designed to induce different affective states: neutral, stress, and amusement. Each participant underwent a protocol that included a baseline (neutral), a stress induction task, and an amusement task.
- Baseline (Neutral) Protocol: Participants were seated or standing at a table and provided with neutral reading materials, such as magazines, for a duration of 20 min.
- Amusement Protocol: Participants viewed eleven humorous video clips, each followed by a brief neutral sequence lasting five seconds. The total duration of the amusement protocol was 392 s.
- Stress Protocol: Participants underwent the Trier Social Stress Test (TSST), a well-established procedure involving public speaking and a mental arithmetic task. Initially, participants delivered a 5-minute speech in front of a three-person panel. Following the speech, the panel instructed participants to count backward from 2023 in decrements of 17. If a mistake was made, participants had to start over. This task also lasted 5 min, making the total duration of the stress protocol 10 min.
- Meditation: Following the amusement and stress conditions, both designed to stimulate the participants, a guided meditation session was conducted. The purpose of this meditation was to calm the participants and return them to a near-neutral emotional state. The meditation involved a controlled breathing exercise, guided by an audio track. Participants were instructed to sit comfortably with their eyes closed while following the audio instructions. The meditation session lasted for 7 min.
- Sensor Placement: Sensors were placed on the chest and wrist of each participant. Specifically, the Empatica E4 wristband was used to collect wrist data, and the RespiBAN was used for chest data.

WESAD recorded data

- Data Types: ECG, EDA, EMG, RESP, TEMP, ACCE were collected.
- Duration: Data collection sessions lasted approximately 2 h per participant, covering the entire protocol.
- Data Quality: Data was collected continuously during the study. Quality control measures included synchronization of the devices and manual inspection to ensure data completeness and accuracy.

In the SWEET study [10], data was collected from 1002 healthy volunteers over five consecutive days to examine daily-life stress in an office worker population. The study utilized a combination of wearable devices and smartphone-based contextual measurements. The first of the two devices is a chest patch intended for measuring ECG and ACCE with a sampling rate of 256 and 32 Hz, respectively. The second wearable device is the imec’s Chillband; this device is worn on the wrist and designed to measure skin conductance (SC), SKT, and ACCE and sampled at 256, 1, and 32 Hz, respectively. Participants also provided self-reported stress levels via Ecological Momentary Assessments (EMAs) triggered 12 times per day by a smartphone application. Stress levels were reported on a five-point Likert scale and later grouped into three classes for analysis: class 1 (no stress) and 2 (light stress) remained unchanged, while classes 3 through 5 were grouped into class 3 (high stress). For binary classification, class 1 was relabeled as class 0 (no stress), and classes 2 through 5 were combined into class 1 (stress). The study also included baseline psychological assessments and contextual data such as location and audio features, with the aim of identifying digital phenotypes for stress detection in daily life. Sensor placement ensured continuous monitoring, with devices worn throughout the day and night, except during specific activities like showering. This study utilized the ECG modality for training and testing on using the same hyperparameters from the WESAD dataset as a proof of concept. Additionally, the approach of [23] was adopted in subject selection and preprocessing, ensuring consistency and reliability in data handling. The datasets made available contained 240 Participants.

SWEET subject information

- Demographics: The dataset included 1002 participants (484 males, 451 females) with an average age of 39.4 ± 9.8 years.
- Inclusion Criteria: Active employees in technology-oriented, banking, and public sector companies.
- Exclusion Criteria: No additional exclusion criteria were applied beyond employment status.

SWEET experimental procedures

- Study Design: The study involved continuous monitoring of physiological signals alongside self-reported stress levels during participants’ daily routines.
- Stress Protocol: Participants were subjected to the Montreal Imaging Stress Task on the first day, which included a stress-inducing arithmetic task.
- Baseline and Contextual Data Collection: Baseline psychological assessments and contextual data (e.g., location, ambient factors) were collected to provide comprehensive data on stressors.

SWEET recorded data

- Data Types: ECG, SC, SKT, ACCE were collected.
- Duration: Data collection spanned 5 days per participant.
- Data Quality: High quality was maintained, with synchronization and quality control procedures ensuring the integrity of the collected data.

Pre-processing

WESAD

- We utilized the BioSPPy library, an open-source tool for biosignal processing. This library provided us with robust and efficient algorithms for the analysis and filtering of Electrocardiogram (ECG), Electrodermal Activity (EDA), and Respiration (RESP) signals. For this study, we used the default parameters provided by the library. For additional information, documentation, and code examples, we recommend visiting the official BioSPPy GitHub repository (https://github.com/PIA-Group/BioSPPy).
- To thoroughly investigate the impact of temporal signal length on stress classification, this study employed a strategic segmentation approach inspired by previous research [15,23]. Window sizes were tested in increments, exploring durations of 60, 120, 210, 300, and 390 ss. Additionally, to examine the effect of overlap, shifts of 10, 20, 30, 60, 120, 210, 300, and 390 ss were applied to each window size appropriately to a total of 31 combinations, including the original unsegmented signal.
- To ensure feature compatibility and improve machine learning model performance, the extracted features were normalized using Z-score. This process involved subtracting the mean and dividing by the standard deviation of each feature.

SWEET

To ensure the accuracy of sensor data, preprocessing steps were adopted as detailed in [22]. ECG readings were evaluated based on heart rate limits (40–180 BPM) and standard deviation criteria, with measurements having a standard deviation below 0.04 discarded. Additionally, a continuous 10-minute window of valid ECG data was required post-cleaning. Data alignment, transformation, and resampling were conducted to standardize the dataset, ultimately producing 9655 samples across 238 users, each representing a window of 60 s with a shift of 60 s.

Feature extraction

Various methods exist for extracting features from physiological signals, including time-domain, frequency-domain, and statistical-based approaches. Statistical features like mean, standard deviation (std), minimum, and maximum have demonstrated satisfactory outcomes in stress classification, whereas flatness and skew have shown relatively modest performance in this application [22].

WESAD

Feature extraction targeted both time-domain and frequency-domain characteristics across modalities. The statistical results were obtained directly from preprocessed signal segments. To analyze frequency patterns, a Fast Fourier Transform (FFT) and power spectral density (PSD) calculations were performed. The same set of statistical features was then derived from the power spectrum to enable comparative analysis across domains (Table 1).

SWEET

Adopting the feature extraction methods described in [22], ECG features were extracted and presented in Table 2.

Feature selection

In our study, feature selection was performed using the Select From Model (SFM) method [24], which employs a Random Forest classifier as a meta-transformer. Specifically, we utilized a Random Forest with $n = 100$ trees to determine feature importance scores, following a methodology analogous to that described in [25]. It is important to note that feature selection was conducted exclusively on the training set, with the test set remaining entirely independent to ensure unbiased and realistic results.

Classifiers and hyperparameter optimization (HPO)

To ensure a comprehensive evaluation and the potential to uncover unexpected relationships; we tested ten different classifiers: Random Forest (RF), Extreme Gradient Boosting (XGB), k-nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree (DT), AdaBoost (AB), Extra Trees (ET), Bagging (BAG), Quadratic Discriminant Analysis (QDA), and Linear Discriminant Analysis (LDA).

As part of the first stage of validation, the dataset was partitioned into a training set comprising 80 % of the data and a test set containing the remaining 20 %. Feature selection was conducted on the training set to identify the most informative features for each modality and segmentation size. Following this, hyperparameter optimization was performed using the GridSearchCV function from the scikit-learn library [26] with 3-CV exclusively on the training data. The optimized hyperparameters and selected feature sets were then validated on the test set. Based on these evaluations, the best-performing model for each modality and segmentation size was selected along with its corresponding optimal hyperparameters. The specific ranges of hyperparameters explored during the Grid Search process are detailed in Table 3.

Table 1
WESAD Extracted Features, Including Time and Frequency Domain Features.

Modality	Domain	N	Feature	Description
ECG	Time	1	HR Mean	Arithmetic average of HR values.
		2	HR Variance	Statistical variance of HR values.
		3	HR std	Standard deviation (std) of HR values.
		4	HR Median	Median of HR values.
		5	HR Maximum	Maximum observed HR value.
		6	HR Minimum	Minimum observed HR value.
		7	HR Q1	First quartile (25th percentile) of HR values.
		8	HR Q3	Third quartile (75th percentile) of HR values.
		9	HR Skewness	Measure of the asymmetry of HR distribution.
		10	HR Kurtosis	Measure of the tailedness of HR distribution.
		11	Mean RR	Average of RR intervals
		12	HR SDNN	Standard deviation of NN intervals
		13	HR RMSSD	Root mean square of successive NN interval differences
	Frequency	1	ECG PSD Mean	Arithmetic average of the ECG PSD values.
		2	ECG PSD Variance	Statistical variance of the ECG PSD values.
		3	ECG PSD std	Standard deviation of the ECG PSD values.
		4	ECG PSD Median	Median of the ECG PSD values.
		5	ECG PSD Maximum	Maximum observed value in the ECG PSD.
		6	ECG PSD Minimum	Minimum observed value in the ECG PSD.
		7	ECG PSD Q1	First quartile (25th percentile) of the ECG PSD values.
		8	ECG PSD Q3	Third quartile (75th percentile) of the ECG PSD values.
		9	ECG PSD Skewness	Measure of the asymmetry of the ECG PSD distribution.
		10	ECG PSD Kurtosis	Measure of the tailedness of the ECG PSD distribution.
EDA	Time	1	EDA Mean	Arithmetic average of EDA values.
		2	EDA Variance	Statistical variance of EDA values.
		3	EDA std	Standard deviation (std) of EDA values.
		4	EDA Median	Median of EDA values.
		5	EDA Maximum	Maximum observed EDA value.
		6	EDA Minimum	Minimum observed EDA value.
		7	EDA Q1	First quartile (25th percentile) of EDA values.
		8	EDA Q3	Third quartile (75th percentile) of EDA values.
		9	EDA Skewness	Measure of the asymmetry of EDA distribution.
		10	EDA Kurtosis	Measure of the tailedness of EDA distribution.
		1	EDA PSD Mean	Arithmetic average of the EDA PSD values.
	Frequency	2	EDA PSD Variance	Statistical variance of the EDA PSD values.
		3	EDA PSD std	Standard deviation of the EDA PSD values.
		4	EDA PSD Median	Median of the EDA PSD values.
		5	EDA PSD Maximum	Maximum observed value in the EDA PSD.
		6	EDA PSD Minimum	Minimum observed value in the EDA PSD.
		7	EDA PSD Q1	First quartile (25th percentile) of the EDA PSD values.
		8	EDA PSD Q3	Third quartile (75th percentile) of the EDA PSD values.
		9	EDA PSD Skewness	Measure of the asymmetry of the EDA PSD distribution.
		10	EDA PSD Kurtosis	Measure of the tailedness of the EDA PSD distribution.
RESP	Time	1	RESP Mean	Arithmetic average of RESP values.
		2	RESP Variance	Statistical variance of RESP values.
		3	RESP std	Standard deviation (std) of RESP values.
		4	RESP Median	Median of RESP values.
		5	RESP Maximum	Maximum observed RESP value.
		6	RESP Minimum	Minimum observed RESP value.
		7	RESP Q1	First quartile (25th percentile) of RESP values.
		8	RESP Q3	Third quartile (75th percentile) of RESP values.
		9	RESP Skewness	Measure of the asymmetry of RESP distribution.
		10	RESP Kurtosis	Measure of the tailedness of RESP distribution.
	Frequency	1	RESP PSD Mean	Arithmetic average of the RESP PSD values.
		2	RESP PSD Variance	Statistical variance of the RESP PSD values.
		3	RESP PSD std	Standard deviation of the RESP PSD values.
		4	RESP PSD Median	Median of the RESP PSD values.
		5	RESP PSD Maximum	Maximum observed value in the RESP PSD.
		6	RESP PSD Minimum	Minimum observed value in the RESP PSD.
		7	RESP PSD Q1	First quartile (25th percentile) of the RESP PSD values.
		8	RESP PSD Q3	Third quartile (75th percentile) of the RESP PSD values.
		9	RESP PSD Skewness	Measure of the asymmetry of the RESP PSD distribution.
		10	RESP PSD Kurtosis	Measure of the tailedness of the RESP PSD distribution.

Table 2
SWEET Extracted Features.

Modality	N	Feature	Description
ECG	1	HR Mean	Arithmetic average of HR values.
	2	ECG SDNN	Standard deviation of RR intervals
	3	ECG RMSSD	Root mean square of successive RR differences
	4	ECG LF	Low-frequency signal (power in the 0.04–0.015 Hz band)
	5	ECG HF	High-frequency signal (power in the 0.15–0.4 Hz band)
	6	ECG LFHF	Ratio of low and high frequency

Table 3
Range of hyperparameters for grid search per classifier.

Classifier	N	Hyperparameter	Range
RF	1	max_depth	[None, 10, 20, 30]
	2	max_features	[’sqrt’, ’log2’]
	3	min_samples_leaf	[1, 2, 4]
	4	min_samples_split	[2, 5, 10, 15]
	5	n_estimators	[50, 100, 200, 300]
XGB	1	learning_rate	[0.01, 0.1, 0.2]
	2	max_depth	[5, 6, 7, 8]
	3	min_samples_split	[2, 3, 4, 5, 6, 7, 10]
	4	n_estimators	[50, 100, 200, 300]
	5	subsample	[0.4, 0.6, 0.8, 1.0]
KNN	1	leaf_size	[5, 10, 20, 30, 40]
	2	n_neighbors	[7, 10, 15, 20]
	3	p	[1, 2]
	4	weights	[’uniform’, ’distance’]
LR	1	C	[0.1, 1, 10, 100]
	2	max_iter	[1500, 2000, 3000]
	3	solver	[’liblinear’, ’lbfgs’, ’sag’, ’newton-cg’]
DT	1	max_depth	[None, 10, 20, 30]
	2	max_features	[’sqrt’, ’log2’]
	3	min_samples_leaf	[1, 2, 4]
	4	min_samples_split	[2, 5, 10, 15, 20]
AB	1	learning_rate	[0.2, 0.5, 0.6, 0.7, 0.8, 0.9]
	2	n_estimators	[50, 100, 200, 300]
ET	1	max_depth	[None, 10, 20, 30, 40, 50, 60]
	2	max_features	[’sqrt’, ’log2’]
	3	min_samples_leaf	[1, 2, 3, 4]
	4	min_samples_split	[2, 5, 10]
	5	n_estimators	[50, 100, 200, 300]
BAG	1	bootstrap	[True, False]
	2	max_features	[0.3, 0.4, 0.5, 1.0]
	3	max_samples	[0.3, 0.4, 0.5, 1.0]
	4	n_estimators	[50, 100, 200, 300]
QDA	1	reg_param	[0.0, 0.1, 0.5, 1.0]
LDA	1	shrinkage	[None, ’auto’, 0.1, 0.5, 0.9]
	2	solver	[’lsqr’, ’eigen’]

Class distribution balancing

To address this issue, SMOTE (Synthetic Minority Oversampling Technique) was applied to the training portion of each cross-validation split. SMOTE is a resampling method frequently employed in machine learning to mitigate class imbalances in datasets, particularly in classification tasks. This technique helps to reduce the bias of the classifier towards the class with a larger number of samples [22].

Evaluation

The best-performing optimized model for each modality was utilized. To ensure a robust evaluation and mitigate potential performance variance due to data splits, K-fold cross-validation (K-CV) with K values ranging from 3 to 10 was employed [27]. This approach provides a more reliable estimate of performance compared to a single train/test split, as it reduces the risk of overfitting and helps assess the model’s generalization to unseen data. Ensemble methods, specifically majority voting (MV) and weighted averaging (WA), were applied to the outputs of ECG, EDA, and RESP. This investigation aimed to determine potential performance gains from a multi-modal approach compared to evaluating each modality independently. Accuracy (ACC) quantifies the proportion of correct predictions, while Precision (P) measures the accuracy of identifying positive labels correctly. Recall (R) indicates the percentage of

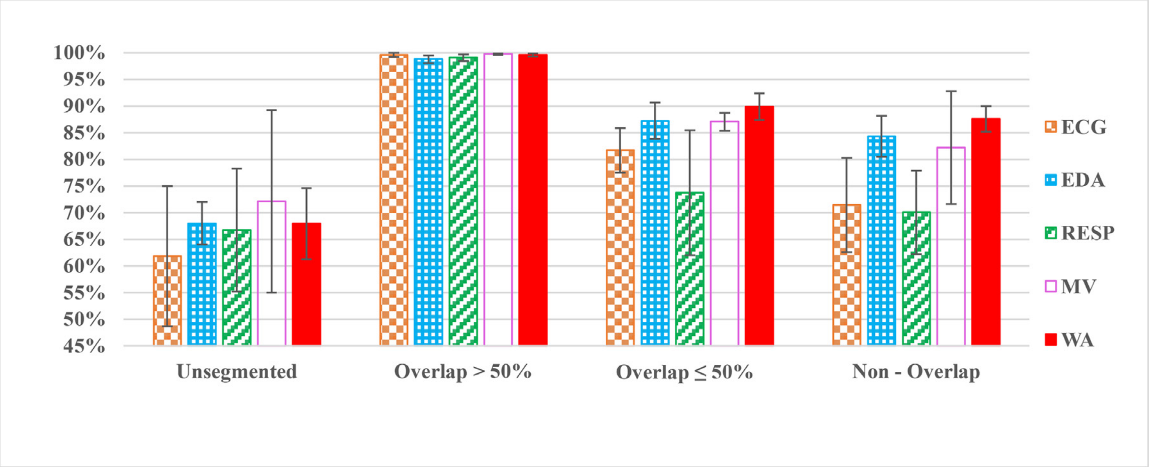

Fig. 2. Best WESAD binary classification ACC results based on time domain features.

actual positive cases the model successfully identifies. The F-measure (F1) score, the harmonic mean of precision and recall, provides a single metric balancing these two aspects. Additionally, the Area Under the Receiver Operating Characteristic Curve (AUC) was used to evaluate the model's performance, reflecting its ability to distinguish between positive and negative classes. These five evaluation metrics were calculated for each fold and averaged across K-CV runs, offering a comprehensive assessment of model performance. This constitutes the second stage of validation, refining the model's robustness before its final evaluation. The equations are defined below.

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

$$\mathrm{F1} = 2 \times \frac{P \times R}{P + R} \tag{4}$$

Further verification was conducted by applying the optimized hyperparameters of the best-performing classifier for the ECG modality to the SWEET dataset, marking the third stage of validation. This step ensured the generalizability and robustness of the findings across a different dataset, validating the model's ability to perform consistently in real-world, unconstrained settings. By testing the model on the SWEET dataset, we confirmed that the performance achieved during the earlier stages of validation (using the WESAD dataset) could be replicated in a more diverse, real-life context, thus reinforcing the reliability of the developed stress detection model.

Method validation

The best time domain performance for the ECG modality was achieved using a 300-second window with a 50 % overlap, yielding an accuracy of 99.63 %±0.37 % with an optimized BAG classifier for binary classification. For multi-class classification, the highest accuracy of 98.36 %±0.94 % was obtained with a 390-second window and 50 % overlap using an optimized XGB classifier. For the EDA modality, the highest binary classification accuracy of 98.78 %±0.73 % was achieved with a 300-second window and 50 % overlap using an ET classifier, while the multi-class classification accuracy peaked at 98.13 %±0.75 % using a 390-second window with 50 % overlap and an optimized ET classifier. For the RESP modality, the best binary classification accuracy of 99.10 %±0.64 % was obtained with a 390-second window and 50 % overlap using an XGB classifier, and the highest multi-class accuracy of 97.68 %±1.08 % was achieved with the same segmentation strategy. These results are summarized in Table 4 and visualized in Figs. 2 and 3.

In terms of ensemble methods, the MV ensemble approach achieved its best performance for binary classification with an optimized XGB, RF, and LDA classifier for ECG, EDA, and RESP, respectively, resulting in an accuracy of 99.78 %±0.15 % using a 210-second window with 50 % overlap. For multi-class classification, the MV ensemble method yielded the highest accuracy of 99.38 %±0.31 % with a 390-second window and 50 % overlap using an optimized XGB, ET, and XGB classifier for ECG, EDA, and RESP, respectively. For the WA ensemble method, the peak binary classification accuracy of 99.60 %±0.25 % was achieved using a 120-second window with 50 % overlap and optimized XGB, ET, and XGB classifiers for ECG, EDA, and RESP. The highest multi-class classification accuracy of 99.61 %±0.32 % was achieved using a 210-second window with 50 % overlap and optimized ET, ET, and XGB classifiers for ECG,

Table 4
Best results based on time domain features.

Modality	Type	Segmentation (Window_Shift)	Model	K-CV	ACC±std (%)	*P*±std (%)	*R*±std (%)	F1±std (%)
ECG	Binary	Unsegmented	XGB	8	61.81±13.17	52.29±30.51	50.00±26.73	49.26±25.03
		300_10	BAG	9	99.63±0.37	99.65±0.56	99.72±0.34	99.68±0.32
		390_210	ET	4	81.70±4.18	83.61±6.87	83.25±7.70	83.11±3.78
		210_210	BAG	8	71.43±8.82	76.52±7.91	72.82±16.40	73.54±9.95
	Multi	Unsegmented	BAG	5	44.00±3.65	55.99±14.20	44.00±3.65	43.63±4.87
		390_10	XGB	10	98.36±0.94	98.44±0.86	98.36±0.94	98.36±0.95
		390_210	ET	4	65.63±6.87	68.63±5.85	65.63±6.87	65.35±7.14
		300_300	KNN	4	56.89±9.37	65.88±10.51	56.89±9.37	57.30±9.50
EDA	Binary	Unsegmented	RF	3	68.00±4.00	62.50±4.17	62.73±15.02	61.94±8.59
		300_20	ET	9	98.78±0.73	99.59±0.62	98.32±1.55	98.94±0.65
		120_60	RF	6	87.21±3.40	87.95±4.09	87.82±5.18	87.79±3.38
		60_60	ET	6	84.29±3.86	85.65±3.69	83.55±6.11	84.49±4.17
	Multi	Unsegmented	XGB	6	36.22±11.59	51.82±16.18	36.22±11.59	39.54±13.17
		390_10	ET	7	98.13±0.75	98.31±0.71	98.13±0.75	98.16±0.75
		60_30	KNN	8	69.54±3.24	76.51±2.63	69.54±3.24	71.11±3.02
		60_60	ET	8	64.06±5.29	72.53±4.99	64.06±5.29	66.06±5.23
RESP	Binary	Unsegmented	LDA	5	66.67±11.55	61.76±22.71	52.86±20.78	56.31±20.42
		390_10	XGB	8	99.10±0.64	99.04±0.83	99.56±0.84	99.29±0.50
		210_120	KNN	6	73.77±11.73	78.35±11.16	70.51±12.79	74.13±11.87
		300_300	RF	3	70.03±7.84	74.95±16.10	70.49±10.53	71.35±4.51
	Multi	Unsegmented	KNN	7	32.21±14.15	40.58±15.21	32.21±14.15	33.44±13.75
		390_10	XGB	6	97.68±1.08	97.80±1.00	97.68±1.08	97.70±1.06
		210_120	KNN	3	49.30±3.04	62.11±1.75	49.30±3.04	52.18±2.98
		210_210	RF	8	45.83±2.46	52.02±6.36	45.83±2.46	46.20±3.05
All (MV)	Binary	Unsegmented	XGB+RF+LDA	9	72.07±17.13	72.41±25.58	65.74±19.30	67.22±18.37
		210_10	BAG+KNN+BAG	6	99.78±0.15	99.88±0.18	99.71±0.27	99.79±0.13
		120_60	XGB+RF+ET	7	87.06±1.69	88.98±2.31	86.09±3.58	87.45±1.83
		210_210	BAG+KNN+QDA	9	82.16±10.59	93.81±9.79	74.95±18.26	81.61±13.17
	Multi	Unsegmented	BAG+XGB+KNN	4	52.19±12.19	50.61±11.78	52.19±12.19	50.61±11.62
		390_10	XGB+ET+XGB	7	99.38±0.31	99.38±0.31	99.38±0.31	99.38±0.31
		120_60	XGB+KNN+KNN	10	76.10±3.63	76.53±3.73	76.10±3.63	75.53±3.54
		60_60	DT+ET+KNN	7	65.85±4.87	65.96±5.31	65.85±4.87	64.72±4.94
All (WA)	Binary	Unsegmented	XGB+RF+LDA	4	67.91±6.72	70.63±20.96	56.25±12.50	59.71±5.05
		120_10	XGB+ET+XGB	7	99.60±0.25	99.55±0.48	99.70±0.32	99.62±0.24
		60_30	RF+ET+RF	8	89.92±2.50	89.46±2.65	91.21±4.02	90.27±2.44
		60_60	ET+ET+XGB	8	87.57±2.41	86.93±2.95	89.38±3.03	88.10±2.29
	Multi	Unsegmented	BAG+XGB+KNN	7	46.75±21.04	58.18±15.19	46.75±21.04	47.65±18.91
		210_10	ET+ET+XGB	9	99.61±0.32	99.62±0.32	99.61±0.32	99.61±0.32
		120_60	XGB+KNN+KNN	10	78.54±4.08	81.52±3.93	78.54±4.08	78.94±3.97
		60_60	DT+ET+KNN	7	67.21±2.89	71.28±2.95	67.21±2.89	68.21±2.69

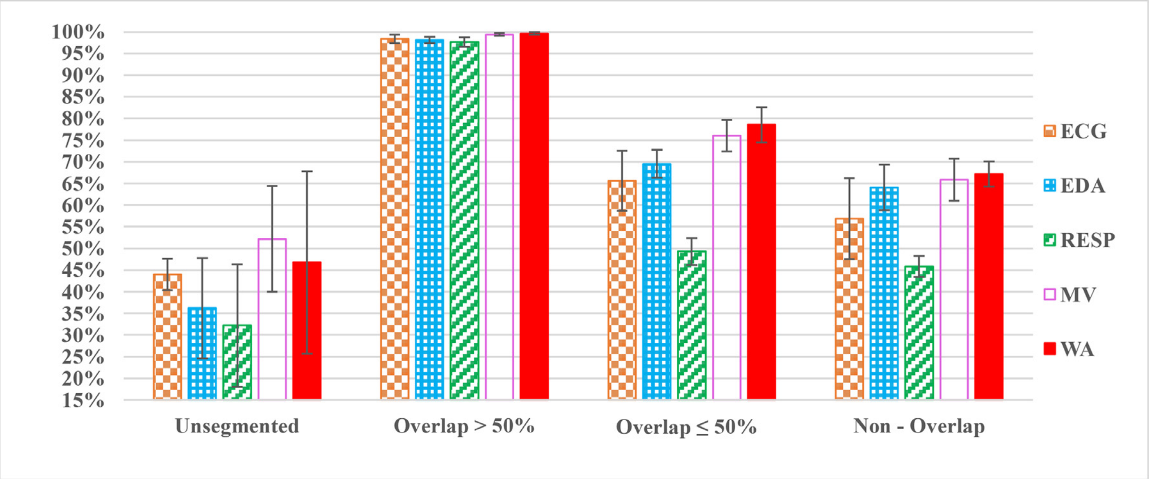

Fig. 3. Best WESAD multi-class classification ACC results based on time domain features.

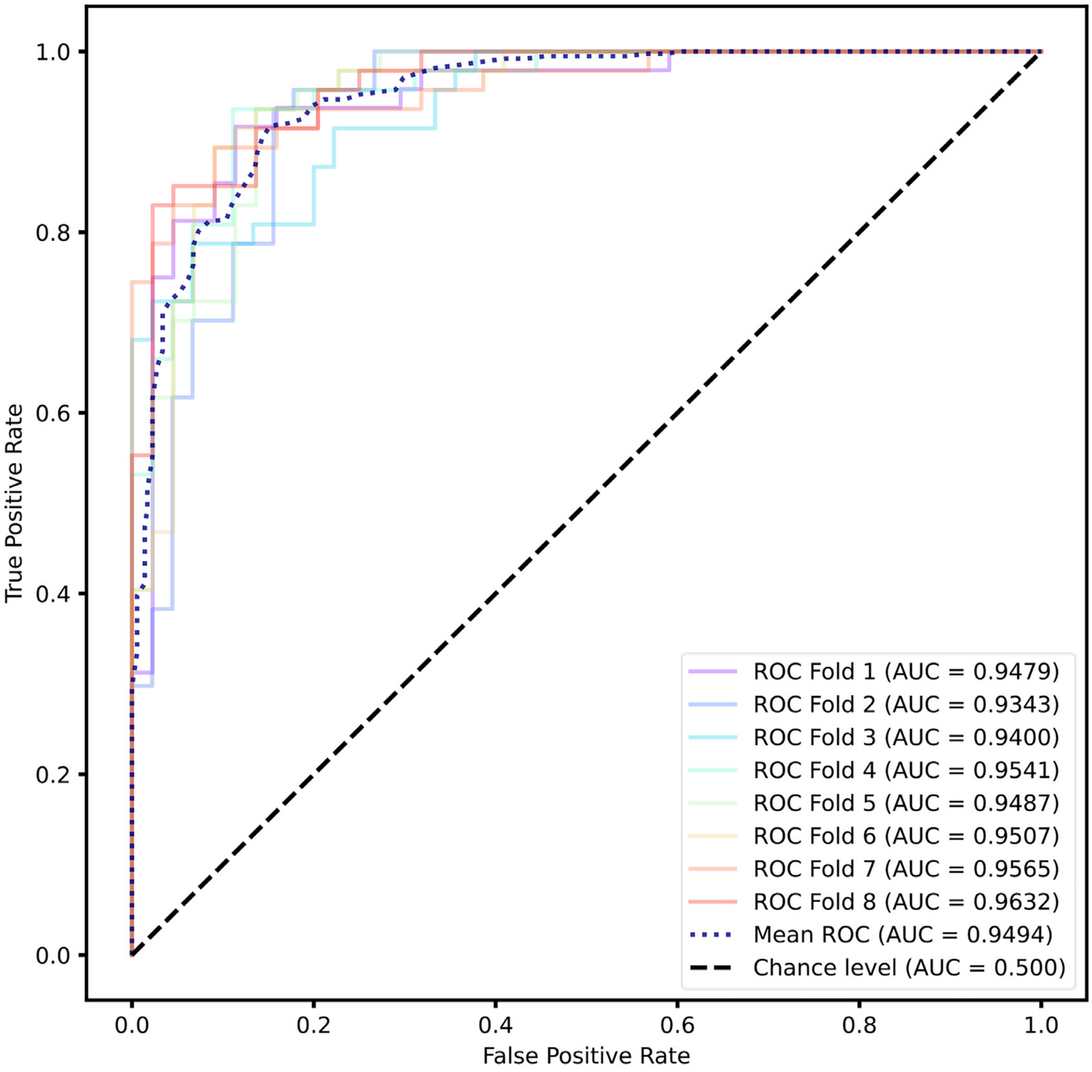

Fig. 4. Binary WESAD classification WA 60_60 ROC curve based on time domain features.

EDA, and RESP. Additionally, the ROC curve for the best non-overlap binary WESAD classification (WA 60_60) based on time-domain features is depicted in Fig. 4. These ensemble results are detailed in Table 4 and further visualized in Figs. 2, 3.

The best frequency domain performance for the ECG modality was achieved with a 390-second window and over 50 % overlap, yielding an accuracy of 96.66 %±1.27 % with an optimized KNN classifier for binary classification. For multi-class classification, the highest accuracy of 95.53 %±1.27 % was obtained using the same window and overlap, with an optimized BAG classifier. In the EDA modality, the highest binary classification accuracy of 94.57 %±1.11 % was achieved with a 390-second window and over 50 % overlap, using an optimized XGB classifier, while the best multi-class classification accuracy of 80.09 %±2.54 % was obtained with a 390-second window and over 50 % overlap using an optimized XGB classifier. For the RESP modality, the highest binary classification accuracy of 88.75 %±1.77 % was achieved with a 390-second window and over 50 % overlap using an optimized ET classifier, and the best multi-class classification accuracy of 75.11 %±3.29 % was obtained with a 390-second window and over 50 % overlap using an optimized ET classifier. These results are summarized in Table 5 and visualized in Figs. 5 and 6.

Regarding ensemble methods, the MV ensemble approach achieved the highest binary classification accuracy of 97.68 %±0.48 % with a 390-second window and over 50 % overlap, using optimized KNN, XGB, and ET classifiers for ECG, EDA, and RESP, respectively. For multi-class classification, the highest accuracy of 93.27 %±1.09 % was achieved with the same window and overlap using optimized ET, XGB, and ET classifiers for ECG, EDA, and RESP. In the WA ensemble method, the best binary classification accuracy of 98.13 %±1.07 % was achieved with a 390-second window and over 50 % overlap, using optimized KNN, XGB, and ET classifiers for ECG, EDA, and RESP, respectively. The highest multi-class classification accuracy of 96.89 %±1.47 % was achieved using the same window and overlap with optimized ET, XGB, and ET classifiers. These ensemble results are detailed in Table 5 and further

Table 5
Best results based on frequency domain features.

Modality	Type	Segmentation (Window_Shift)	Model	K-CV	ACC±std (%)	*P*±std (%)	*R*±std (%)	F1±std (%)
ECG	Binary	Unsegmented	LDA	4	70.76±6.35	70.36±9.84	53.13±11.97	60.35±11.17
		390_10	KNN	10	96.66±1.27	97.78±1.29	96.98±2.06	97.36±1.02
		210_120	RF	10	74.14±9.65	77.43±9.97	74.42±11.49	75.50±8.80
		300_300	ET	9	67.72±12.96	67.59±11.50	70.24±17.43	68.66±14.14
	Multi	Unsegmented	BAG	9	39.04±12.94	47.51±13.40	39.04±12.94	41.65±11.01
		390_10	ET	6	95.53±1.27	95.69±1.22	95.53±1.27	95.55±1.27
		120_60	DT	10	53.90±8.49	60.32±7.50	53.90±8.49	55.22±8.17
		120_120	ET	6	54.77±8.66	61.19±8.91	54.77±8.66	56.16±9.20
EDA	Binary	Unsegmented	KNN	7	78.96±11.35	76.43±14.64	72.14±16.80	74.05±15.45
		390_10	XGB	7	94.57±1.11	95.58±0.94	95.91±1.13	95.74±0.87
		60_30	ET	5	83.17±2.50	85.46±2.78	81.18±5.65	83.15±2.87
		120_120	RF	6	80.05±3.45	80.23±4.46	81.22±7.78	80.45±3.68
	Multi	Unsegmented	LR	9	64.35±14.86	61.72±23.18	64.35±14.86	61.32±17.85
		390_10	XGB	6	80.09±2.54	85.81±1.93	80.09±2.54	81.01±2.30
		60_30	BAG	4	64.05±2.11	73.36±2.90	64.05±2.11	66.53±2.18
		120_120	BAG	5	59.28±3.18	68.51±3.68	59.28±3.18	61.72±3.83
RESP	Binary	Unsegmented	LR	4	69.30±6.79	67.50±8.66	56.25±21.65	59.46±13.66
		390_10	ET	10	88.75±1.77	93.36±2.36	88.72±3.08	90.93±1.53
		210_120	LR	5	67.93±6.05	74.75±7.75	61.53±5.61	67.39±5.77
		210_210	XGB	6	69.05±6.65	74.26±9.83	71.60±6.43	72.42±5.12
	Multi	Unsegmented	LDA	8	40.00±9.22	46.69±12.72	40.00±9.22	41.09±9.53
		390_10	ET	7	75.11±3.29	78.84±2.48	75.11±3.29	75.79±3.05
		390_210	KNN	10	42.03±11.17	47.35±14.80	42.03±11.17	40.49±10.51
		390_390	ET	4	44.67±16.81	49.38±20.07	44.67±16.81	46.33±18.42
All (MV)	Binary	Unsegmented	LDA+KNN+LR	3	76.00±4.00	76.85±1.60	62.73±7.77	68.95±5.34
		390_10	KNN+XGB+ET	7	97.68±0.48	97.91±1.18	98.49±1.52	98.18±0.39
		210_120	RF+DT+LR	10	83.79±6.09	87.01±7.52	82.63±8.85	84.44±6.28
		210_210	DT+KNN+XGB	9	78.69±10.50	81.04±9.71	82.32±9.57	81.45±8.62
	Multi	Unsegmented	BAG+LR+LDA	9	59.26±14.81	50.69±18.35	59.26±14.81	53.00±15.20
		390_10	ET+XGB+ET	8	93.27±1.09	93.81±1.16	93.27±1.09	93.26±1.11
		60_30	ET+BAG+KNN	7	65.93±1.94	66.69±2.18	65.93±1.94	65.80±1.99
		60_60	RF+XGB+BAG	7	62.15±3.05	62.95±2.58	62.15±3.05	61.67±3.07
All (WA)	Binary	Unsegmented	LDA+KNN+LR	6	75.85±12.00	74.44±17.21	68.89±20.07	70.15±15.43
		390_10	KNN+XGB+ET	10	98.13±1.07	98.25±1.35	98.85±0.73	98.54±0.83
		60_30	RF+ET+AB	10	85.89±3.10	83.93±3.42	89.85±4.09	86.73±2.96
		60_60	BAG+BAG+AB	10	83.60±2.84	82.92±4.13	86.17±3.65	84.42±2.53
	Multi	Unsegmented	BAG+LR+LDA	8	51.94±12.64	54.89±16.70	51.94±12.64	51.26±12.42
		390_10	ET+XGB+ET	7	96.89±1.47	97.02±1.36	96.89±1.47	96.90±1.46
		60_30	ET+BAG+KNN	8	72.67±4.19	77.97±3.60	72.67±4.19	73.83±3.98
		120_120	ET+BAG+ET	8	69.39±6.60	73.99±5.96	69.39±6.60	70.00±6.66

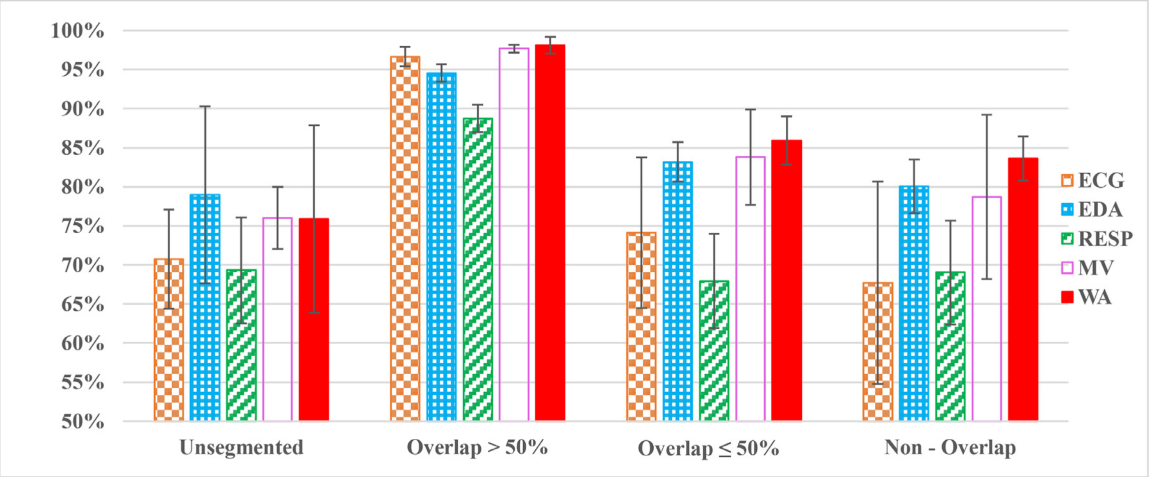

Fig. 5. Best WESAD binary classification ACC results based on frequency domain features.

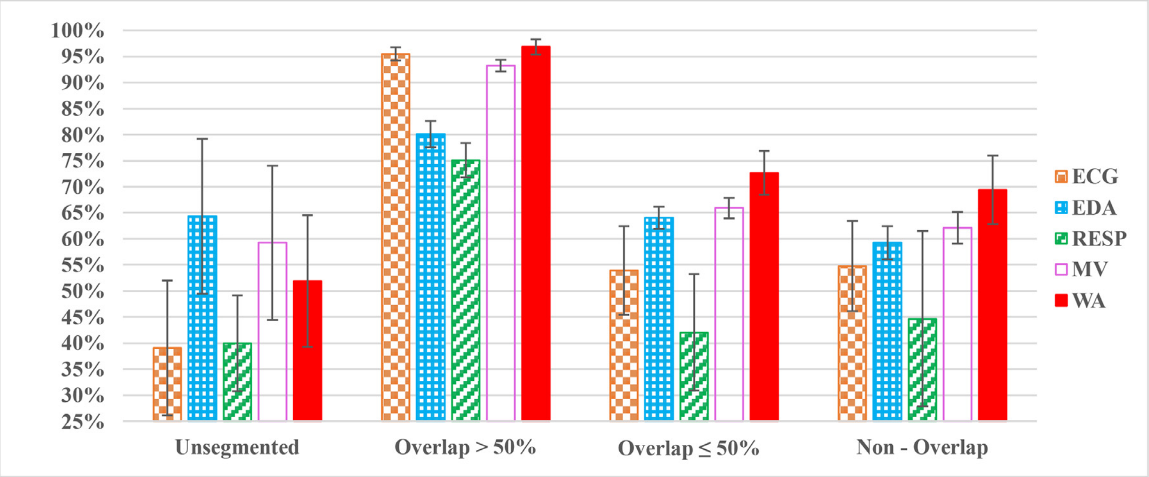

Fig. 6. Best WESAD multi-class classification ACC results based on frequency domain features.

Table 6
Comparative evaluation metrics of WESAD and SWEET datasets using the 60_60 WESAD configuration.

Modality	Dataset	Type	Segmentation (Window_Shift)	Model	K-CV	ACC±std (%)	*P*±std (%)	*R*±std (%)	F1±std (%)
ECG	WESAD	Binary	60_60	ET	5	68.99±4.00	70.14±3.73	69.22±5.49	69.64±4.39
		Multi		DT	7	47.41±3.49	53.74±4.62	47.41±3.49	48.83±3.55
	SWEET	Binary		ET	5	73.72±1.54	52.95±2.23	64.72±3.08	58.24±2.53
		Multi		DT	7	57.09±1.94	68.33±1.47	57.09±1.94	60.50±1.75

Table 7
Comparison with related studies.

Paper References	Dataset	Modality	Validation	Model	ACC%	Our Proposal
Zhu et al. (2023) [28]	CLAS UTD VerBIO WESAD	EDA Non-overlap	LOSO	SVM RF SVM RF	68.5 73.1 92.9 86.5	EDA 60_60 6-CV ET 84.29±3.86
Adarsh et al. (2024) [29]	WESAD SWELL	ECG Overlap>50 %	5-CV	GCN	97.75 94.48	ECG 300_10 9-CV BAG 99.63±0.37
Schmidt et al. (2018) [15]	WESAD	RESP Overlap>50 %	LOSO	LDA	88.09	RESP 390_10 8-CV XGB 99.10±0.64
Rashid et al. (2023) [30]	WESAD	ECG+EDA+RESP Overlap>50 % ACCE+ECG+EDA Overlap>50 %	LOSO	AB	81.62 86.37	MV 210_10 6-CV BAG+KNN+BAG 99.78±0.15

visualized in Figs. 5 and 6. Additionally, the ROC curve for the best non-overlap binary WESAD classification (WA 60_60) based on frequency-domain features is shown in Fig. 7.

Given that the SWEET dataset is segmented into 60_60 intervals, we applied the optimal 60_60 configuration identified from the WESAD dataset, including the best classifier and its associated hyperparameters. Since ECG is the sole modality shared between WESAD and SWEET, it was used as the basis for comparison. The results, detailed in Table 6, illustrate that the WESAD-derived configuration generalizes effectively, as evidenced by the superior performance on the SWEET dataset compared to WESAD, as depicted in Fig. 8. Furthermore, the ROC curve for the SWEET results, shown in Fig. 9, further supports the effectiveness and generalizability of the chosen configuration.

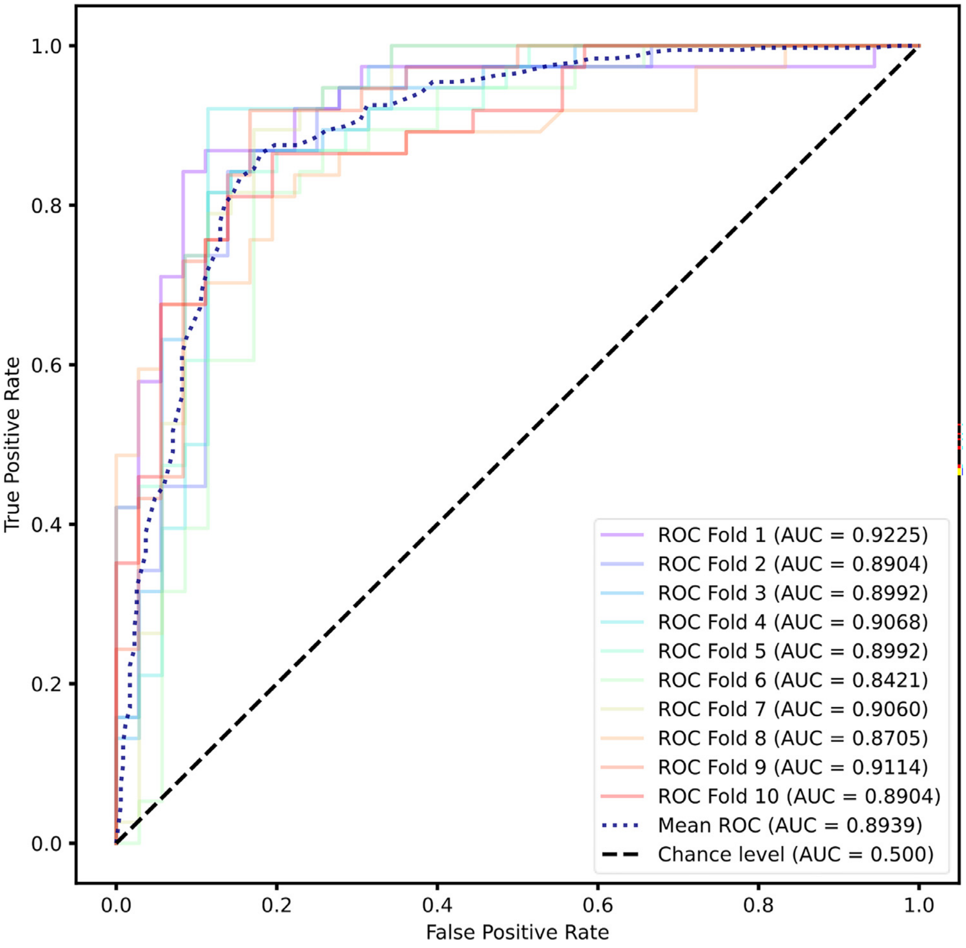

Fig. 7. Binary WESAD classification WA 60_60 ROC curve based on frequency domain features.

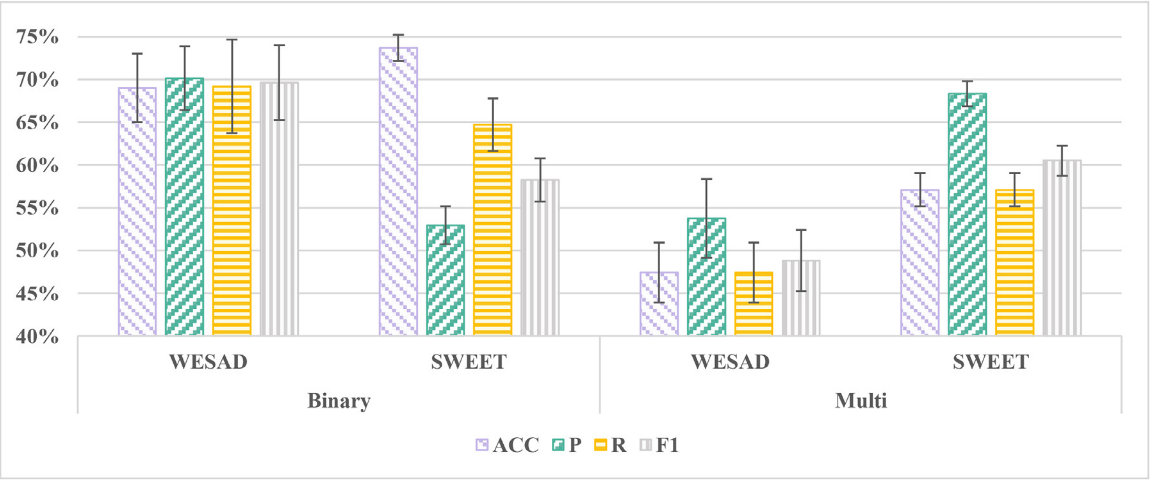

Fig. 8. Comparison of evaluation metrics between the ECG modality from the WESAD and SWEET datasets using the 60_60 configuration derived from WESAD.

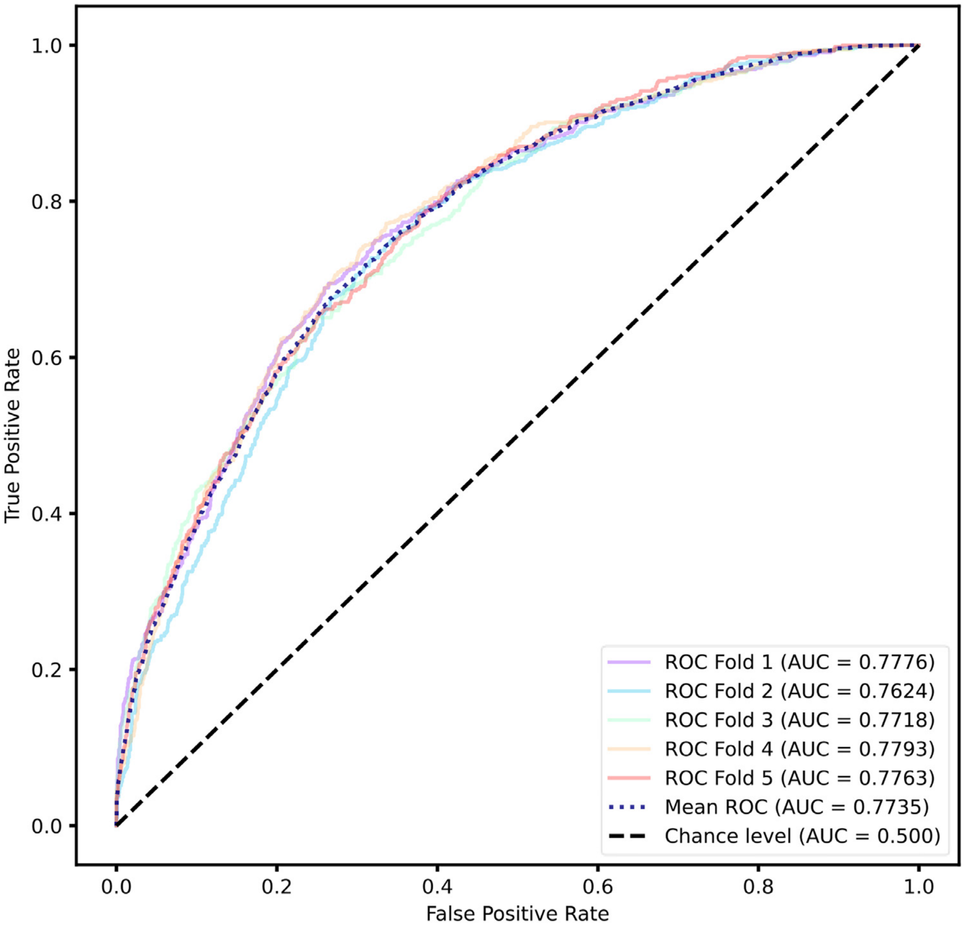

Fig. 9. Binary SWEET classification ECG 60_60 ROC curve.

Table 7 juxtaposes our research findings with relevant studies, delineating comparisons based on the utilization of individual modalities, namely, ECG, EDA, and RESP. Each of our employed modalities is juxtaposed against corresponding studies employing singular modalities. Additionally, our ensemble methodology is compared against studies integrating all three modalities simultaneously.

Zhu et al. (2023) [28] conducted research on binary stress classification utilizing exclusively the EDA modality from four distinct datasets: CLAS, UTD, VerBIO, and WESAD. The primary focus is on the WESAD dataset, which our study employed. In their investigation, an accuracy of 86.5 % was achieved by utilizing segmentation without overlap, employing a 30-second window size, and employing the RF classifier with LOSO cross-validation. Conversely, our study attained a slightly lower accuracy of 84.29 %±3.86 % under similar settings, exclusively utilizing the EDA modality, employing segmentation without overlap, employing a 60-second window size, and utilizing the ET classifier with 6-CV.

Adarsh et al. (2024) [29] investigated binary stress classification by leveraging the ECG modality from two distinct datasets: SWELL and WESAD. The primary focus is on the WESAD dataset, which our study employed. In the research conducted by the authors, an accuracy of 97.75 % was achieved through segmentation with >50 % overlap, utilizing a window size of 5 ss with a 0.25-second shift, and employing graph convolutional Networks (GCN) with 5-CV. Conversely, the investigation conducted in our study yielded a higher accuracy of 99.63 %±0.37 % under analogous conditions, exclusively employing the ECG modality, employing segmentation with >50 % overlap, utilizing a window size of 300 ss with a 10-second shift, and employing the BAG classifier with 9-CV.

Schmidt et al. (2018) [15] explored binary stress classification by utilizing the Respiration (RESP) modality from the WESAD dataset, which was also employed in our study. In the investigation conducted by Schmidt et al., an accuracy of 88.09 % was achieved through segmentation with >50 % overlap, employing a window size of 60 ss with a 0.25-second shift, and utilizing LDA with LOSO cross-validation. Conversely, our study attained a higher accuracy of 99.10 %±0.64 % under similar conditions, exclusively utilizing the RESP modality, employing segmentation with >50 % overlap, utilizing a window size of 390 ss with a 10-second shift, and employing the XGB classifier with 8-CV.

Rashid et al. (2023) [30] investigated binary stress classification through a multimodal approach incorporating ECG, EDA, and RESP modalities from the WESAD dataset, which aligns with the dataset utilized in our study. In their investigation, Rashid et al.

Table 8
Commercial wearable devices that provide ECG, EDA, and RESP modalities.

Device Name	Body Location	ECG	EDA	RESP	Additional Features
Fitbit Sense 2	Wrist	Yes	Yes	Yes	SpO2, SKT
Flowtime	Head	Yes	No	No	2-channel brainwave
Movesense	Chest	Yes	No	No	Motion measurement
Prana	Weist	No	No	Yes	Posture
Sentio Solutions Feel Therapeutics	Wrist	Yes	Yes	No	Physical activity, SKT

Table 9
Recommended configurations for optimal performance.

System	Modality	Type	Features Domain	Segmentation (Window_Shift)	Model
Offline	ECG	Binary	Time	210_210	BAG
		Multi	Time	300_300	KNN
	EDA	Binary	Time	60_60	ET
		Multi	Time	60_60	ET
	RESP	Binary	Time	300_300	RF
		Multi	Time	210_210	RF
	All (WA)	Binary	Time	60_60	ET+ET+XGB
		Multi	Frequency	120_120	ET+BAG+ET
Online	ECG	Binary	Time	300_10	BAG
		Multi	Time	390_10	XGB
	EDA	Binary	Time	300_20	ET
		Multi	Time	390_10	ET
	RESP	Binary	Time	390_10	XGB
		Multi	Time	390_10	XGB
	All (MV)	Binary	Time	210_10	BAG+KNN+BAG
	All (WA)	Multi	Time	210_10	ET+ET+XGB

achieved an accuracy of 81.62 % by employing segmentation with >50 % overlap, utilizing a window size of 60 ss with a 5-second shift, and employing AB with LOSO cross-validation. Conversely, our study achieved a notably higher accuracy of 99.78 %±0.15 % under analogous conditions, employing a multimodal WA ensemble of ECG, EDA, and RESP modalities. Our methodology involved segmentation with >50 % overlap, utilizing a window size of 210 ss with a 10-second shift, and employing BAG, KNN, and BAG classifiers for ECG, EDA, and RESP, respectively, with 6-CV.

We utilized three different modalities: ECG, EDA, and RESP. Additionally, we employed two ensemble methods that simultaneously integrated these modalities. Our investigation into commercially available wearable devices that offer these modalities is summarized in Table 8, derived from Taskasaplidis et al. (2024) [31].

Based on our findings, we recommend the configurations presented in Table 9 for optimal performance in real-world applications tailored to the system type and available modalities.

Limitations

None.

Ethics statements

Not Applicable.

Declaration of competing interest

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

CRediT authorship contribution statement

Basil A. Darwish: Writing – original draft, Methodology, Formal analysis, Data curation. **Shafiq Ul Rehman:** Writing – review & editing, Validation, Methodology, Conceptualization. **Ibrahim Sadek:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Nancy M. Salem:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Ghada Kareem:** Writing – review & editing, Validation, Methodology, Conceptualization. **Lamees N. Mahmoud:** Writing – review & editing, Validation, Methodology, Conceptualization.

Data availability

Two datasets were used, one publicly available and the other is confidential.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] A.F.A. Mentis, D. Lee, P. Roussos, Applications of artificial intelligence−machine learning for detection of stress: a critical overview, Mol. Psychiatry (2023) 1–13 Apr. 2023, doi:10.1038/s41380-023-02047-6.

[2] S. Sharma, G. Singh, M. Sharma, A comprehensive review and analysis of supervised-learning and soft computing techniques for stress diagnosis in humans, Comput. Biol. Med. 134 (Jul. 2021) 104450, doi:10.1016/J.COMPBIOMED.2021.104450.

[3] R. Li, Z. Liu, Stress detection using deep neural networks, BMC. Med. Inform. Decis. Mak. 20 (11) (Dec. 2020) 1–10, doi:10.1186/S12911-020-01299-4/TABLES/5.

[4] A. Arsalan, M. Majid, Human stress classification during public speaking using physiological signals, Comput. Biol. Med. 133 (Jun. 2021) 104377, doi:10.1016/J.COMPBIOMED.2021.104377.

[5] S. Cohen, D. Janicki-Deverts, Who's stressed? Distributions of psychological stress in the United States in probability samples from 1983, 2006, and 20091, J. Appl. Soc. Psychol. 42 (6) (Jun. 2012) 1320–1334, doi:10.1111/J.1559-1816.2012.00900.X.

[6] M. Kivimäki, A. Steptoe, Effects of stress on the development and progression of cardiovascular disease, Nat. Rev. Cardiol. 15 (4) (Apr. 2018) 215–229, doi:10.1038/NRCARDIO.2017.189.

[7] J. Wang, et al., The application of machine learning techniques in posttraumatic stress disorder: a systematic review and meta-analysis, NPJ. Digit. Med. 7 (1) (May 2024) 1–13 2024 7:1, doi:10.1038/s41746-024-01117-5.

[8] N.K. Iyortsuun, S.H. Kim, M. Jhon, H.J. Yang, S. Pant, A review of machine learning and deep learning approaches on mental health diagnosis, Healthcare 11 (3) (Feb. 2023), doi:10.3390/HEALTHCARE11030285.

[9] W.N. Price, I.G. Cohen, Privacy in the age of medical big data, Nat. Med. 25 (1) (Jan. 2019) 37–43 2019 25:1, doi:10.1038/s41591-018-0272-7.

[10] E. Smets, et al., Large-scale wearable data reveal digital phenotypes for daily-life stress detection, NPJ. Digit. Med. 1 (1) (Dec. 2018) 1–10 2018 1:1[dataset], doi:10.1038/s41746-018-0074-9.

[11] A.B.R. Shatte, D.M. Hutchinson, S.J. Teague, Machine learning in mental health: a scoping review of methods and applications, Psychol. Med. 49 (9) (Jul. 2019) 1426–1448, doi:10.1017/S0033291719000151.

[12] E.J. Topol, High-performance medicine: the convergence of human and artificial intelligence, Nat. Med. 25 (1) (Jan. 2019) 44–56 2019 25:1, doi:10.1038/s41591-018-0300-7.

[13] M. Ghassemi, T. Naumann, P. Schulam, A.L. Beam, I.Y. Chen, R. Ranganath, A review of challenges and opportunities in machine learning for health, AMIa Jt. Summits. Transl. Sci. Proc. 2020 (2020) 191–200 available: http://www.ncbi.nlm.nih.gov/pubmed/32477638 .

[14] E. Smets, et al., Comparison of machine learning techniques for psychophysiological stress detection, Commun. Comput. Inf. Sci. 604 (2016) 13–22, doi:10.1007/978-3-319-32270-4_2/FIGURES/3.

[15] P. Schmidt, A. Reiss, R. Duerichen, K. Van Laerhoven, Introducing WeSAD, a multimodal dataset for wearable stress and affect detection, in: ICMI 2018 - Proceedings of the 2018 International Conference on Multimodal Interaction, Association for Computing Machinery, Inc, Oct. 2018, pp. 400–408, doi:10.1145/3242969.3242985. [dataset].

[16] Y.S. Can, N. Chalabianloo, D. Ekiz, C. Ersoy, Continuous stress detection using wearable sensors in real life: algorithmic programming contest case study, Sensors 19 (2019) 1849 Pagevol. 19, no. 8, p. 1849, Apr. 2019, doi:10.3390/S19081849.

[17] E.E. Kaczor, B. Chapman, S. Carreiro, P. Indic, J. Stapp, Objective measurement of physician stress in the Emergency department using a wearable sensor, Proc. Annu Hawaii. Int. Conf. Syst. Sci. (2020) 3729 2020, doi:10.24251/hicss.2020.456.

[18] T. Iqbal, et al., A sensitivity analysis of biophysiological responses of stress for wearable sensors in connected health, IEEe Access. 9 (2021) 93567–93579, doi:10.1109/ACCESS.2021.3082423.

[19] A. Greco, et al., Acute stress State classification based on electrodermal activity modeling, IEEe Trans. Affect. Comput. 14 (1) (Jan. 2023) 788–799, doi:10.1109/TAFFC.2021.3055294.

[20] T. Iqbal, A. Elahi, W. Wijns, A. Shahzad, Exploring unsupervised machine learning classification methods for physiological stress detection, Front. Med. Technol. 4 (Mar. 2022) 782756, doi:10.3389/FMEDT.2022.782756/BIBTEX.

[21] R. Kuttala, R. Subramanian, V.R.M. Oruganti, Multimodal hierarchical CNN feature fusion for stress detection, IEEe Access. 11 (2023) 6867–6878, doi:10.1109/ACCESS.2023.3237545.

[22] M. Abd Al-Alim, R. Mubarak, N.M. Salem, I. Sadek, A machine-learning approach for stress detection using wearable sensors in free-living environments, Comput. Biol. Med. 179 (Sep. 2024) 108918, doi:10.1016/J.COMPBIOMED.2024.108918.

[23] M. Albaladejo-González, J.A. Ruipérez-Valiente, F.Gómez Mármol, Evaluating different configurations of machine learning models and their transfer learning capabilities for stress detection using heart rate, J. Ambient. Intell. Humaniz. Comput. 14 (8) (Aug. 2023) 11011–11021, doi:10.1007/S12652-022-04365-Z/TABLES/6.

[24] M. Huljanah, Z. Rustam, S. Utama, T. Siswantining, Feature selection using random forest classifier for predicting prostate cancer, IOP. Conf. Ser. Mater. Sci. Eng. 546 (5) (Jun. 2019) 052031, doi:10.1088/1757-899X/546/5/052031.

[25] R. Richer, et al., Machine learning-based detection of acute psychosocial stress from body posture and movements, Sci. Rep. 14 (1) (Apr. 2024) 1–19 2024 14:1, doi:10.1038/s41598-024-59043-1.

[26] F. Pedregosa Fabianpedregosa, et al., Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (85) (2011) 2825–2830, doi:10.5555/1953048.2078195.

[27] M.A. Aboamer, A.T. Azar, A.S.A. Mohamed, K.J. Bär, S. Berger, K. Wahba, Nonlinear features of heart rate variability in paranoid schizophrenic, Neural Comput. Appl. 25 (7–8) (Dec. 2014) 1535–1555, doi:10.1007/S00521-014-1621-1/TABLES/8.

[28] L. Zhu, et al., Stress detection through wrist-based electrodermal activity monitoring and machine learning, IEEe J. Biomed. Health Inform. 27 (5) (May 2023) 2155–2165, doi:10.1109/JBHI.2023.3239305.

[29] V. Adarsh, G.R. Gangadharan, Mental stress detection from ultra-short heart rate variability using explainable graph convolutional network with network pruning and quantisation, Mach. Learn. (Jan. 2024) 1–28, doi:10.1007/S10994-023-06504-9/TABLES/6.

[30] N. Rashid, T. Mortlock, M.A. Al Faruque, Stress detection using context-aware sensor fusion from wearable devices, IEEe Internet. Things. J. 10 (16) (Aug. 2023) 14114–14127, doi:10.1109/JIOT.2023.3265768.

[31] G. Taskasaplidis, D.A. Fotiadis, P.D. Bamidis, Review of stress detection methods using wearable sensors, IEEe Access. 12 (2024) 38219–38246, doi:10.1109/ACCESS.2024.3373010.