# Multi-Class Stress Detection Using Electrodermal Activity: Evaluation of A Hybrid Model Using UBFC-Phys Dataset

Kawther Alsayed, Hamza Ghandorh, Wael M.S. Yafooz
Department of Computer Science-College of Computer Science and Engineering, Taibah University, Medinah, Saudi Arabia

***Abstract*—Stress is a psychological and physiological response to internal or external pressures or challenges that exceed an individual's ability to cope. In response to these conditions, the human body produces physiological signals that reflect its internal states, often without conscious awareness. Among these signals, electrodermal activity (EDA) is known for its sensitivity to changes in stress levels. Traditional methods for assessing stress, such as questionnaires and self-reports, remain widely used, but they are often influenced by subjectivity and recall bias. This has led to increased interest in objective, data-driven approaches. This research proposes a hybrid stress detection model. The EDA signals were used from the UBFC-Phys dataset, which comprises data from 56 participants under three conditions: rest, moderate stress, and high stress. The signals were preprocessed using filtering, smoothing, and subject-level normalization to reduce inter-individual variability. A set of features was extracted, such as tonic component features. Multiple experiments were conducted on baseline, deep learning, and hybrid models not only for stress detection classification purposes but also for performance evaluation purposes. The proposed model achieved an accuracy of 92.16% in the multi-class classification task. The findings of this research contribute to mental health and well-being by providing an accurate model for stress detection that can be adapted to healthcare, education, and workplace settings, ensuring a healthier and more sustainable future.**

***Keywords*—*Stress detection; physiological signal; Electrodermal Activity (EDA); machine learning; ensemble; deep learning; hybrid model***

## I. Introduction

Stress negatively affects both physical and mental health. It is causative of conditions such as cardiovascular diseases, weakened immunity, anxiety, depression, and sleep disorders [1], [2]. Its effects are not limited to the individual level but extend to productivity and quality of life in general, such as work environments and educational institutions [3], [4]. This highlights the importance of identifying stress early and monitoring it over time to support overall well-being [5]. When a person experiences psychological stress, the autonomic nervous system, particularly the sympathetic system, is triggered. As a result, hormones such as cortisol and adrenaline are released to prepare the body for a fight-or-flight response [6]. These hormones increase blood pressure, heart rate, and breathing rate, and lead to noticeable changes in physiological signals [6]. After the stress ends, the parasympathetic nervous system helps the body recover and return to a normal state.

Recently, more attention has been given to using physiological signals to detect stress directly and continuously. One of the most common signals is electrodermal activity (EDA), a physiological signal that captures changes in skin conductance associated with sweat gland activation by the sympathetic nervous system [7]. Current progress in machine learning (ML) and deep learning (DL) makes it possible to automatically analyze physiological signals and classify stress levels. Various ML models have been applied to learn patterns from extracted features. DL approaches, e.g., Convolutional Neural Networks (CNN) and Transformers, have also been used to capture more complex relationships within the data.

Despite recent progress in stress detection, working with the UBFC-Phys dataset [8] still presents a challenge. This dataset requires careful handling due to the nature of the physiological signals. There are clear differences between individuals, as each person has their own baseline and responds differently to stress. In addition, some patterns, especially between moderate and high stress, can be very similar, which makes classification more challenging. Thus, it is important to use methods that take these differences into account and provide consistent results across all states. The main contributions of this work are:

Addressing inter-individual variability by applying subject-level normalization, which reduces differences in physiological baselines and improves the consistency of stress classification.

Proposed a hybrid model for binary and multiclass stress classification and achieved high performance on the UBFC-Phys dataset.

The remainder of this study is organized as follows: Section II presents a review of the relevant literature. Section III details the materials and methods, including the dataset description, preprocessing procedures, feature extraction process, data partitioning strategy, proposed model architecture, and evaluation metrics. Section IV presents the experiments and results, followed by a comparison with previous work. Section V discusses the findings. Section VI concludes the study.

## II. Literature Review

Recent studies have shown increasing interest in stress detection using physiological signals, as these signals provide more objective indicators. Many publicly available datasets have been widely used in stress detection research to develop, train, and evaluate various ML and DL models using physiological signals. These datasets provide important benchmarks for comparing stress detection models in different environments and with different signals. The UBFC-Phys dataset was used in this research. Other datasets include SWELL-KW [9], which is designed to evaluate workplace stress under time-pressure and email interruption conditions, and WESAD [10], which provides a multimodal dataset that involves physiological signals recorded from chest- and wrist-worn devices during different conditions. Other datasets that contribute to stress detection research comprise SRAD [11], Louisiana [12], TILES-2018 [13], and MAUS [14], which provide data related to driving stress, hospital work environments, workplace stress, and mental workload.

Physiological signals are used for stress detection with the support of wearable devices and AI models. Different ML approaches have been applied in this area. Sabour et al. presented the UBFC-Phys dataset and used EDA and BVP signals to study social stress, showing their usefulness in this field [8]. Vos et al. developed an ensemble model using data from multiple sources and showed good generalization for stress prediction [15]. Hasanpoor et al. used PPG signals from the UBFC-Phys dataset with a CNN-MLP model and reported an accuracy of 82%, showing that a single signal can be used for stress detection [16].

Zhang et al. used a DL model based on attention to combine BVP and EDA signals and capture the relationship between them [17]. Arsalan et al. [18] applied several supervised learning models with EEG, GSR/EDA, and PPG signals, and reported that the MLP model achieved the highest accuracy. Al Hasan et al. [19] used PPG, ECG, and GSR from the MAUS dataset and reported good performance using the Random Forest (RF) model after hyperparameter tuning. Mathur et al. [20] investigated stress detection among nurses. In their experiment, they used wearable sensor data and achieved high performance with Decision Tree classifiers.

WESAD and other datasets have been used to compare ML and DL models for stress detection. Kar et al. used a GRU-RNN model with EDA, accelerometer, and temperature signals and achieved high performance in binary classification [21]. Gil-Martin et al. applied CNN models using data from wearable sensors and showed that DL can learn useful patterns from physiological signals [22]. Liu et al. used a Transformer-LSTM model with PPG signals and reported strong results in stress classification [23]. Other work explored CNN-LSTM models and multimodal signals to improve performance, as combining different signals helps capture more information about the body's response to stress [24]. These studies show that DL models can handle complex patterns, especially when enough data is available.

Many physiological signals were used for stress detection, such as electrodermal activity (EDA), blood volume pulse (BVP), electrocardiogram (ECG), photoplethysmography (PPG), and heart rate variability (HRV) [25]. Electrodermal activity (EDA), on the other hand, is a practical option because it is directly related to sympathetic nervous system activity and can be collected using wearable devices [7]. Despite these improvements, distinguishing between different stress levels is still challenging, especially when the patterns are very similar. Based on this, the current study focuses on using Electrodermal activity (EDA) signals from the UBFC-Phys dataset and proposes a hybrid model for both binary and multiclass stress classification.

## III. Methodology

### A. Dataset

Data were obtained from a publicly accessible dataset named UBFC-Phys [8]. The dataset contains physiological signals collected using the E4 wristband EDA and BVP together with video recordings from 56 participants during three tasks performed under different stress conditions. The experiment included three tasks representing different stress levels: a resting task (T1) used as the baseline condition, a speech task (T2), and an arithmetic task (T3). Each task included 56 recordings, resulting in a total of 168 signals. This work focused on the EDA signals. For each subject, the signals were recorded separately for each task, which made it easier to distinguish between rest and stress states.

Fig. 1 shows that EDA signals increased with higher stress levels across the tasks. The signal was stable during rest (T1), slightly higher during the speaking task (T2), and highest during the arithmetic task (T3), showing that EDA can reflect changes in stress levels.

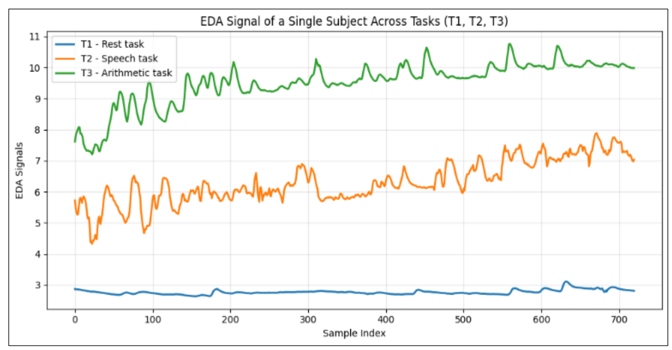


Fig. 1. EDA signal of Subject 13 across the three tasks: rest (T1), speech (T2), and arithmetic (T3). The signal amplitude increases from the resting condition to the more demanding tasks, reflecting higher physiological arousal under stress.

### B. Data Preprocessing

The EDA signals were recorded at a sampling rate of 8 Hz. Several preprocessing steps were applied to reduce noise, enhance the quality of the signal, and prepare the data. First, the initial two seconds of each signal were removed, as this part often contained unstable values due to sensor contact at the beginning of the recording.

A fourth-order Butterworth low-pass filter was applied to attenuate high-frequency artifacts with a 1 Hz cutoff frequency, while keeping the main signal pattern unchanged. A Savitzky-Golay filter was used to smooth and reduce small fluctuations of the signal without affecting the overall shape.

The dataset showed variability between participants in their physiological responses. To address this, Z-score normalization was applied for each subject to reduce these differences and allow the model to focus on relative changes in the EDA signal rather than absolute values. The mean and standard deviation (SD) were calculated from the resting condition (T1) only and subsequently used to normalize all recordings for each participant across different tasks.

$$x\_norm = \frac{(x - \mu\_baseline)}{\sigma_{baseline} + \varepsilon} \tag{1}$$

where x was the EDA value, μ_baseline and σ_baseline were the mean and standard deviation of the resting signal (T1), and ε is a small constant.

### C. Feature Extraction

For each EDA signal, 23 features were extracted to represent different signal characteristics. The extracted features were directly used for model classification. The tonic and phasic components of each EDA signal were extracted using NeuroKit2 [26]. Six features were extracted from the tonic component to describe the baseline level, variability, and overall trend of the signal. Similarly, six features were derived from the phasic component to capture rapid changes related to sympathetic activity.

Additionally, seven statistical features were computed directly from the preprocessed signal to represent its overall distribution and variability. Two Hjorth [27] parameters were also included to describe the signal dynamics and complexity. The zero-crossing rate of the first derivative was used to reflect local variations in the signal, and low-frequency spectral power was calculated to capture the dominant physiological patterns. A summary of extracted features is shown Table I.

TABLE I. FEATURES DERIVED FROM EDA SIGNALS [8]

| Feature | Description |
|---|---|
| scl_mean | Mean of skin conductance level (SCL) |
| scl_std | Tonic component standard deviation |
| scl_range | Range (max − min) of the tonic component |
| scl_slope | Linear trend (slope) of the tonic component over time |
| scl_median | Median value of the tonic component |
| scl_cv | Coefficient of variation of the tonic component |
| scr_mean | Mean absolute value of the phasic (SCR) component |
| scr_std | Standard deviation of the phasic component |
| scr_max | Maximum value of the phasic component |
| scr_auc | Area under the positive phasic component curve |
| scr_energy | Total energy of the phasic component |
| scr_count | Number of detected SCR peaks |
| raw_mean | Mean of the raw EDA signal |
| raw_std | Standard deviation of the raw EDA signal |
| raw_var | Variance of the raw EDA signal |
| raw_range | Range (max - min) of the raw EDA signal |
| raw_skew | Skewness of the raw EDA signal distribution |
| raw_kurt | Kurtosis of the raw EDA signal distribution |
| raw_rms | Root mean square of the raw EDA signal |
| hjorth_mobility | Rate of change of the signal relative to its variance |
| hjorth_complexity | Rate of change of the signal's frequency content |
| zcr | Rate of sign changes in the first derivative of the signal |
| lf_power | Power spectral density in the low-frequency band (0-1 Hz) |

### D. Data Splitting

The dataset was partitioned into training and testing sets using a 70:30 ratio to enable reliable model evaluation. A subject-independent split was applied using GroupShuffleSplit with a fixed random state (42), where all recordings from each participant were assigned to either the training or testing set to avoid data leakage. Within the training set, hyperparameter tuning was conducted using five-fold cross-validation. GroupKFold was used for binary classification, and StratifiedGroupKFold was applied for multiclass tasks. GridSearchCV was used to select the best hyperparameters, and the final model was evaluated on the unseen data (test set).

### E. Experimental Setup

The experimental setup in this work was organized into three experiments.

*1) Binary classification using ML models:* The first experiment evaluated baseline ML models for binary stress classification using EDA-based features. T1 and T2 from the UBFC-Phys dataset were used, where T1 represented the non-stress state, and T2 represented the stress state. The experiment was conducted under different preprocessing settings, without smoothing and normalization and with smoothing and normalization, to examine their effect on classification performance. Fig. 2 shows an overview of the first experiment for binary stress classification. The configurations used in the experiment are summarized in Table II.

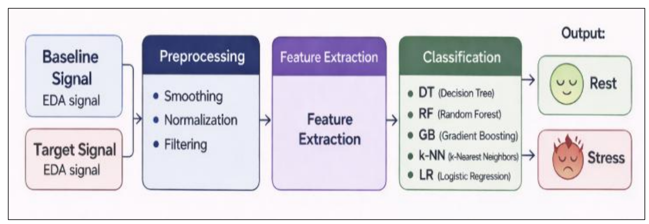


Fig. 2. Overview of the binary Stress Detection Pipeline. It starts with EDA signals (Baseline and Target), then preprocesses them. After that, features are extracted and fed into ML models to classify the stress.

TABLE II. HYPERPARAMETER SETTINGS FOR BASELINE ML MODELS IN BINARY CLASSIFICATION

| Model | HyperParameter |
|---|---|
| GB | n_estimators = 100, 150, 200; learning_rate = 0.01, 0.05, 0.1; max_depth = 2, 3 |
| RF | n_estimators = 200, 300, 500; max_depth = 5, 10, 15, 20, None; min_samples_split = 2, 5; min_samples_leaf = 1, 2 |
| K-NN | n_neighbors = 3, 5, 7, 9, 11; weights = uniform, distance; metric = euclidean, manhattan, minkowski |
| LR | C = 0.001, 0.01, 0.1, 0.5, 1, 5, 10; max_iter = 2000 |
| DT | max_depth = 3, 5, 10, 15, None; min_samples_split = 2, 5, 10; criterion = gini, entropy |

*2) Multiclass classification using ML and DL models*: The second experiment evaluated multi-class stress classification using both ML and DL models. Three tasks from the dataset were used, including T1 for rest, T2 for moderate stress, and T3 for high stress. The same preprocessing steps, feature extraction process, and data splitting strategy used in the first experiment were also applied (Fig. 3). The configurations used in the experiment are summarized in Table III.

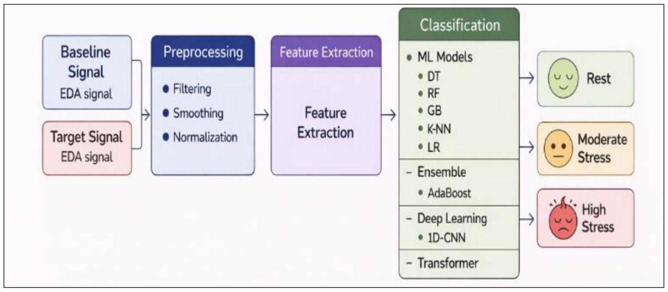


Fig. 3. Overview of the Multi-Level Stress Detection Pipeline: From raw EDA signals, the data preprocessing, and then feature extraction. Different models then use the extracted features to predict three stress levels.

TABLE III. HYPERPARAMETER SETTINGS FOR MODELS IN MULTICLASS CLASSIFICATION

| Model | HyperParameter |
|---|---|
| GB | n_estimators = 100, 150, 200; learning_rate = 0.01, 0.05, 0.1; max_depth = 2, 3 |
| RF | n_estimators = 200, 300; max_depth = 10, 20, None |
| K-NN | n_neighbors = 3, 5, 7, 9; weights = uniform, distance |
| LR | C = 0.01, 0.1, 1, 10; solver = lbfgs |
| DT | max_depth = 3, 5, 10, None; criterion = gini, entropy |
| AdaBoost | base_estimator = Decision Tree (max_depth = 2), SAMME; n_estimators = 50, 100, 200; learning_rate = 0.01, 0.1, 1.0 |
| 1D-CNN | Filters = 64 / 128; FC Hidden = 64; Dropout = 0.4; Learning Rate = 0.0005; Epochs = 150; Batch Size = 16 |
| Feature Transformer | d_model = 64; Attention Heads = 8; Encoder Layers = 2; FC Hidden = 64; Dropout = 0.3; Learning Rate = 0.0005; Epochs = 200; Batch Size = 16; Noise Std = 0.02 |

*3) Proposed hybrid model for multiclass classification:* The proposed hybrid model combined five ML classifiers, Gradient Boosting (GB), K-Nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF), using a weighted soft voting strategy. GB was given the highest weight due to its strong capability for sequential error correction. RF provided robustness. KNN contributed by capturing local data structures, LR introduced a linear decision perspective, and DT added model diversity with low complexity. Fig. 4 shows an overview of the proposed model.

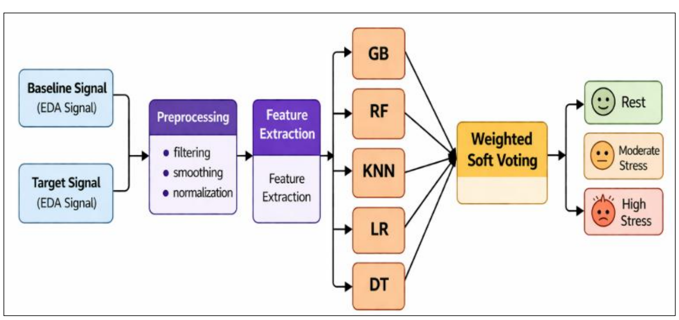


Fig. 4. Overview of the hybrid Stress Detection Pipeline for Multiclass: It starts with EDA signals, followed by preprocessing, feature extraction, and the hybrid model predicts three stress levels: Rest, Moderate Stress, and High Stress.

The configurations used in the experiment are summarized in Table IV.

TABLE IV. THE PROPOSED HYBRID MODEL CONFIGURATION

| Model | HyperParameter | Weight |
|---|---|---|
| GB | n_estimators = 150, learning_rate = 0.01, max_depth = 2 | 3 |
| RF | n_estimators = 300, max_depth = 10, min_samples_split = 5 | 2 |
| K-NN | n_neighbors = 7, weights = distance | 2 |
| LR | C = 10, solver = lbfgs | 1 |
| DT | criterion = gini, max_depth = 3 | 1 |

## F. Performance Evaluation

Several evaluation metrics were employed to assess various aspects of model performance. The models were evaluated on the test set using widely adopted performance measures, including accuracy, recall, precision, F1-score, ROC curves, and the confusion matrix. The performance metrics utilized in this work are summarized in Table V.

TABLE V. EVALUATION METRICS UTILIZED FOR PERFORMANCE ASSESSMENT

| Metric | Description |
|---|---|
| PR | $PR = \frac{TP}{TP + FP} \times 100$ |
| RE | $RE = \frac{TP}{TP + FN} \times 100$ |
| AC | $AC = \frac{\mathrm{TP} + \mathrm{TN}}{Total\ Samples} \times 100$ |
| F1 | $F1 - score = 2 \times \frac{PR\ x\ RE}{PR + RE}$ |
| CM | Provides class-wise information on correct classifications and misclassification patterns. |
| ROC | Shows how the model performance changes across different decision thresholds |

# IV. RESULTS

In the first experiment, baseline ML models were evaluated for binary stress classification without applying smoothing and normalization. Table VI shows the performance results of the models using EDA-based features for classifying between stress and non-stress states.

TABLE VI. THE RESULTS OF THE BASELINE EXPERIMENT (WITHOUT SMOOTHING AND NORMALIZATION)

| Model | AC | PR | RE | F1 |
|---|---|---|---|---|
| **KNN** | **91.18** | **85.00** | **100** | **91.89** |
| GB | 88.24 | 80.95 | 100 | 89.47 |
| LR | 85.29 | 83.33 | 88.24 | 85.71 |
| RF | 82.35 | 76.19 | 94.12 | 84.21 |
| DT | 82.35 | 82.35 | 82.35 | 82.35 |

Fig. 5 and Fig. 6 show the confusion matrices for the selected models in the first experiment. Both GB and KNN achieved high recall for stress, with some false positives where rest samples were misclassified as stress. KNN showed slightly fewer misclassifications compared to GB.

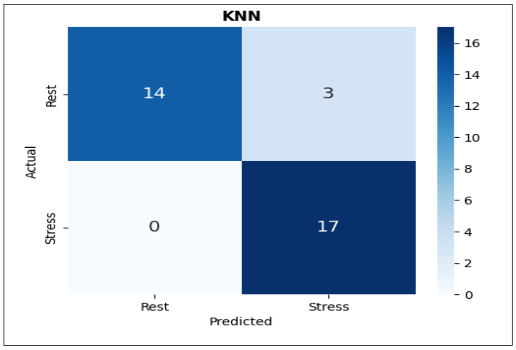


Fig. 5. CM of binary classification (no smooth/no norm) for KNN model.

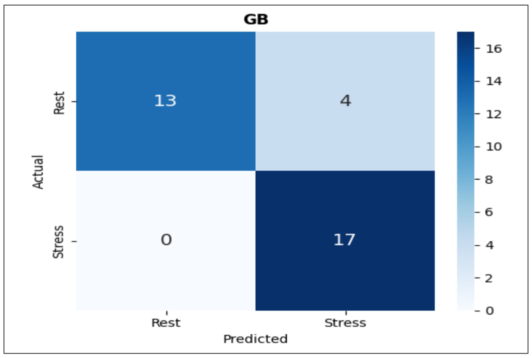


Fig. 6. CM of binary classification (no smooth/no norm) for GB model.

The ROC curves in Fig. 7 and 8 below show that KNN achieved the highest AUC (0.976), which indicated the best separation between the two classes. GB also had a high AUC (0.934), showing good performance.

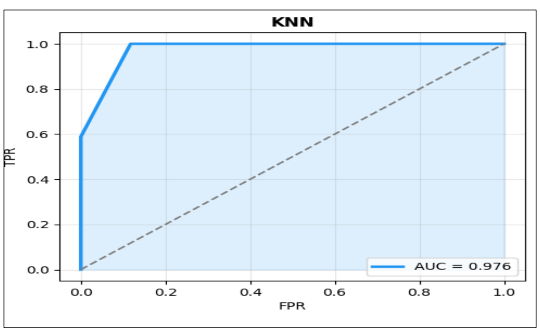


Fig. 7. ROC curve of binary classification (no smooth/no norm) for KNN model.

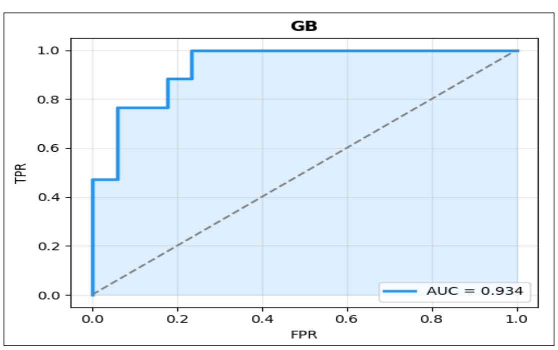


Fig. 8. ROC curve of binary classification (no smooth/no norm) for GB model.

Table VII presents the results after applying smoothing and normalization. The preprocessing steps improved the performance of most models compared to the previous setting. DT achieved perfect results. These results show that the preprocessing pipeline helped the models better distinguish between the classes and improved generalization across different subjects. The high performance may also be related to the relatively small dataset size and the simpler nature of the binary classification task.

TABLE VII. THE RESULTS OF THE BINARY CLASSIFICATION (WITH SMOOTHING AND NORMALIZATION)

| Model | AC | PR | RE | F1 |
|---|---|---|---|---|
| DT | **100** | **100** | **100** | **100** |
| RF | 97.06 | 100 | 94.12 | 96.97 |
| GB | 97.06 | 100 | 94.12 | 96.97 |
| KNN | 91.18 | 85.00 | 100 | 91.89 |
| LR | 88.24 | 88.24 | 88.24 | 88.24 |

Fig. 9 and 10 show the confusion matrices for DT and RF in binary classification using preprocessed EDA signals. DT correctly classified all samples without errors, while RF also showed strong performance with only one misclassification.

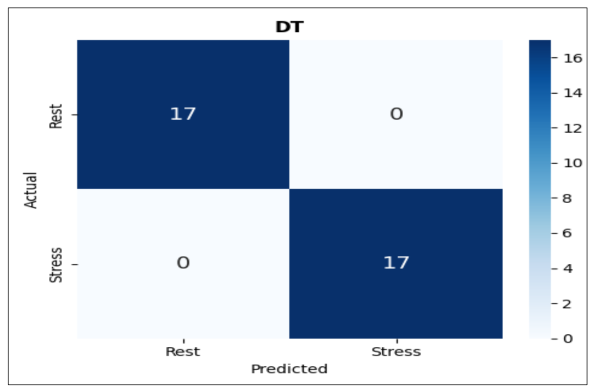


Fig. 9. CM of binary classification (with smooth and norm) for DT model.

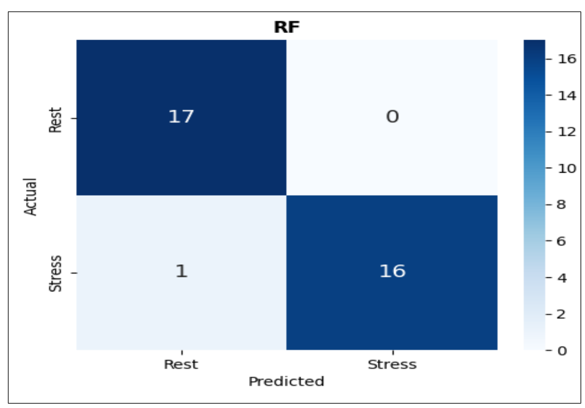


Fig. 10. CM of binary classification (with smooth and norm) for RF model.

Fig. 11 and 12 show the ROC curves for both DT and RF, which achieved perfect AUC values of 1.000, indicating a clear separation between the rest and stress classes.

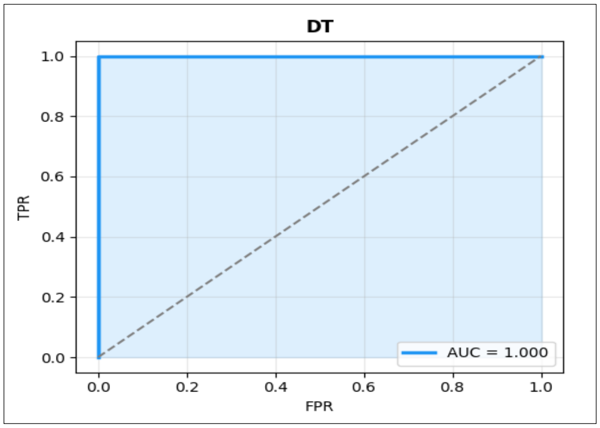


Fig. 11. ROC curve of binary classification (with smooth and norm) for DT model.

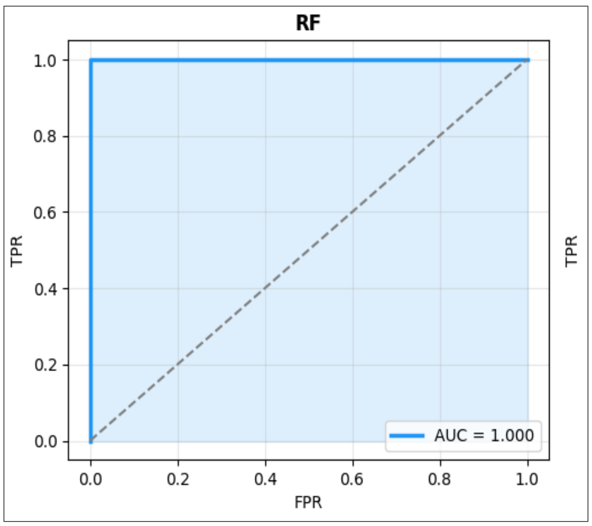


Fig. 12. ROC curve of binary classification (with smooth and norm) for RF model.

The second experiment evaluated multi-class stress classification using both ML and DL models. The three-level stress classification results are presented in Table VIII. The best performance was achieved by GB, while comparable results were obtained by RF, KNN, AdaBoost, and DT. Lower performance was observed for LR and Transformer, whereas the weakest results were recorded by 1D-CNN.

TABLE VIII. RESULTS OF THE MULTICLASS CLASSIFICATION

| Model | AC | PR | RE | F1 |
|---|---|---|---|---|
| GB | **90.20** | **90.28** | **90.20** | **90.19** |
| RF | 88.24 | 88.24 | 88.24 | 88.24 |
| KNN | 88.24 | 88.24 | 88.24 | 88.24 |
| AdaBoost | 88.24 | 88.24 | 88.24 | 88.24 |
| DT | 88.24 | 88.24 | 88.24 | 88.24 |
| LR | 86.27 | 86.73 | 86.27 | 86.36 |
| Transformer | 82.35 | 82.41 | 82.35 | 82.34 |
| 1D-CNN | 52.94 | 47.15 | 52.94 | 44.23 |

The relatively low performance of the 1D-CNN might be due to the small size of the dataset and the subject-independent split, whereby the number of samples available for deep feature learning was reduced. The three classes were well balanced, but the low-frequency EDA and the subtle differences between moderate and high stress could have affected the discriminative representation that the 1D-CNN was able to learn. Therefore, the poor performance of the lower 1D-CNN can be attributed to the model architecture used and less training data.

The confusion matrices for the selected models are shown in Fig. 13 and 14 below show the confusion matrices for the selected models. The GB model correctly classified all the rest of the samples and achieved strong performance across stress levels. The Feature Transformer also correctly identified all the remaining samples but showed more confusion between stress levels compared to GB.

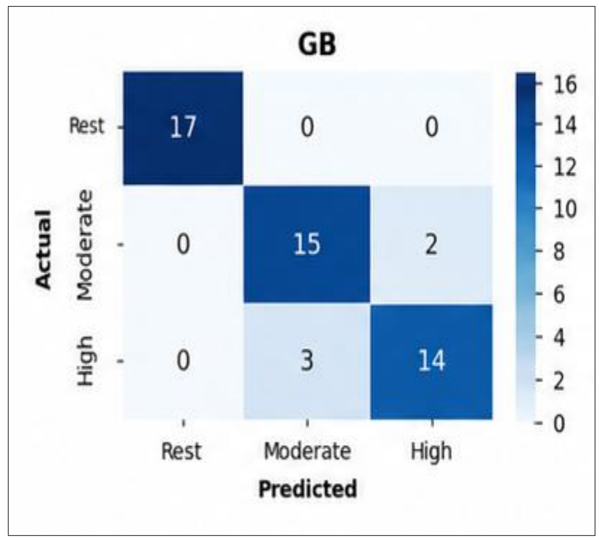


Fig. 13. CM of multiclass classification for the GB model.

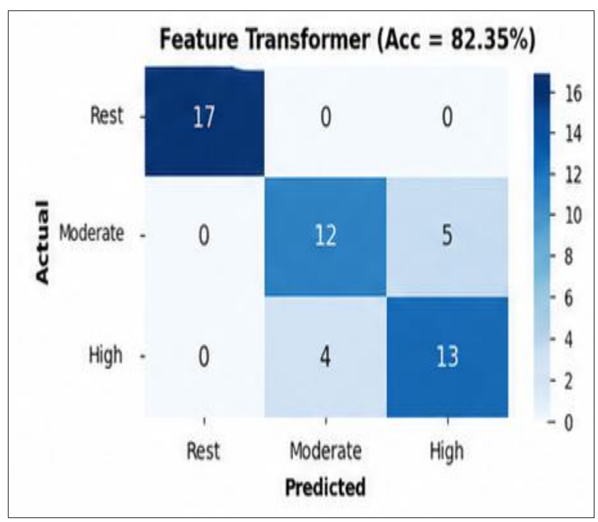


Fig. 14. CM of multiclass classification for Transformer.

Fig. 15 and 16 below show the ROC curves for the selected models. Both models achieved strong performance with clear separation for the rest of the classes.

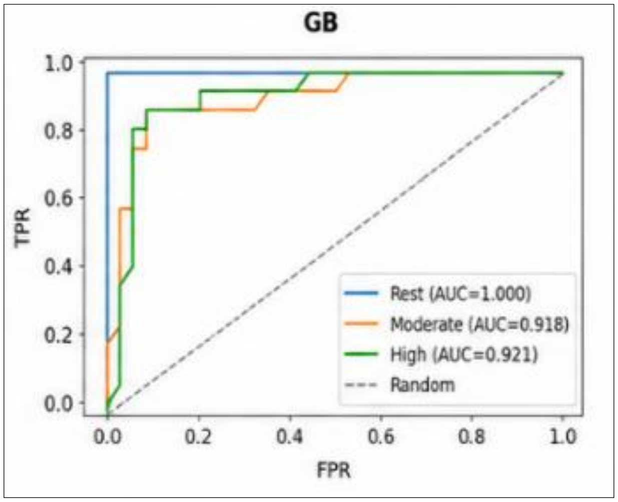


Fig. 15. ROC curve of multiclass classification for GB model.

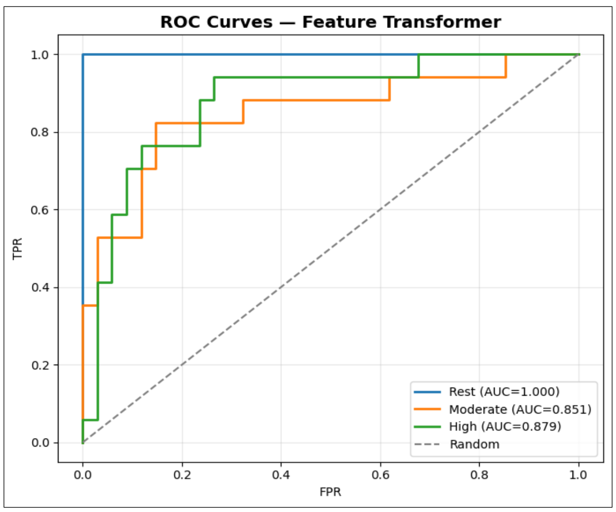


Fig. 16. ROC curve of multiclass classification for Transformer.

The final experiment evaluated the proposed hybrid model. Table IX presents the performance results of the proposed model. The proposed hybrid model outperformed all other evaluated models in terms of overall performance.

TABLE IX. THE RESULTS OF PROPOSED MODEL

| Model | AC | PR | RE | F1 |
|---|---|---|---|---|
| Proposed Hybrid Model | 92.16 | 92.16 | 92.16 | 92.16 |

The confusion matrix in Fig. 17 shows that all the rest of the samples were correctly classified by the proposed hybrid model. Most moderate and high stress samples were also classified correctly, with only a few errors between them. To analyze the contribution of extracted EDA features, the proposed hybrid model was implemented with feature group ablation analysis. One feature group was eliminated in each experiment, and the other features were kept. The results are presented in Table X.

The full 23-feature set achieved the highest accuracy of 92.16%. Removing the phasic features resulted in the greatest drop in performance, indicating their importance in capturing stress-related sympathetic responses. Removing tonic and dynamic/spectral features also shows a reduction, further indicating that different groups of EDA features provide complementary information for multi-class stress classification.

TABLE X. FEATURE-GROUP ABLATION ANALYSIS OF THE PROPOSED HYBRID MODEL

| Feature setting | Features | AC |
|---|---|---|
| Full feature set | **23** | **92.16** |
| Without tonic features | **17** | **88.24** |
| Without phasic features | **17** | **86.27** |
| Without statistical features | **16** | **90.20** |
| Without dynamic/spectral features | **19** | **88.24** |

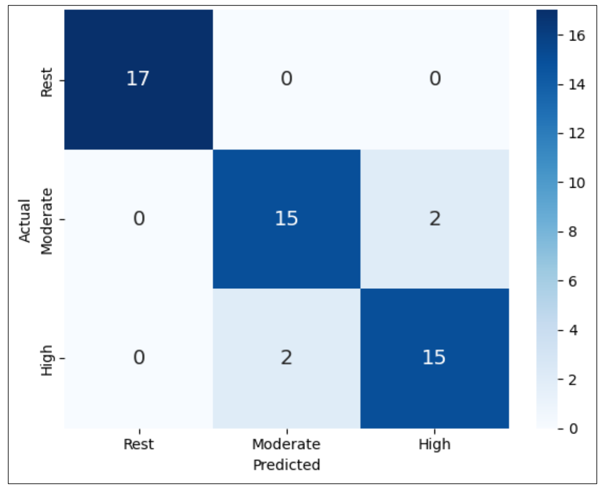


Fig. 17. CM of multiclass classification for the proposed hybrid model.

The ROC curves in Fig. 18 for the proposed hybrid model show a perfect AUC of 1.000 for the rest class. The high and moderate stress classes recorded AUCs of 0.946 and 0.943, respectively, confirming strong discrimination ability across all stress levels.

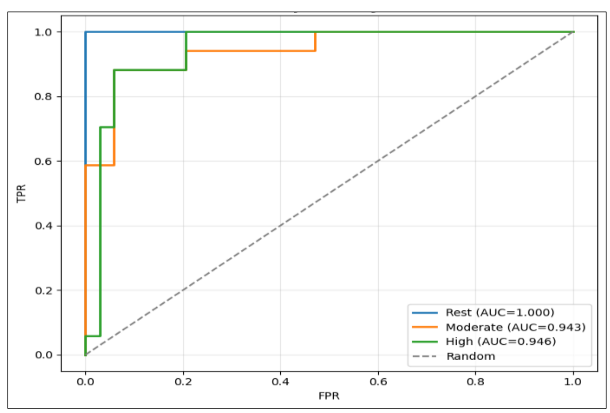


Fig. 18. ROC curve of multiclass classification for proposed hybrid model.

The result of the proposed model was compared with previous stress detection studies that used EDA signals on the UBFC-Phys dataset. Table XI compares the proposed model with EDA-based baseline models evaluated on the UBFC-Phys dataset.

TABLE XI. COMPARISON OF THE PROPOSED MODEL WITH EXISTING MULTICLASS STRESS DETECTION MODELS ON THE UBFC-PHYS DATASET

| Reference | Method | Signal Used | AC |
|---|---|---|---|
| Zhang et al. [17] | DL (Attention) | EDA | 48.1% |
| Sabour et al. [8] | SVM | EDA | 63.60% |
| Valerio and Mahmoud [28] | SVM | EDA | 69.12% |
| **Proposed Model** | **Weighted soft-voting** | **EDA** | **92.16%** |

The comparison shows that the proposed hybrid model achieved the highest accuracy among the EDA-based models evaluated on the UBFC-Phys dataset. The performance of the proposed model was compared with the published SVM baseline and individual EDA-based ML/DL baselines for the multi-class stress classification task and outperformed both. This means that the proposed model is capable of capturing stress-related patterns without using any other signals except EDA signals.

## V. DISCUSSION

The results of this work provide useful insight for computer science in ML, signal processing, and affective computing. The proposed model shows that a single physiological signal, EDA, could achieve high AC in stress classification when it was properly processed using steps such as filtering, smoothing, subject-level normalization, and feature extraction. This means that it was possible to build an effective stress detection system without relying on multiple sensing methods. EDA signals provided meaningful information for distinguishing between different stress levels. Separating closely related stress states could be challenging, especially between the speech task (T2) and the arithmetic task (T3), as they show similar patterns. In contrast, the binary task, which uses T1 and T2, was easier to distinguish, leading to higher performance. The proposed model achieved a high multiclass AC of 92.16%, showing that EDA could be effectively used as a standalone signal for stress detection. The improvement in performance highlights the importance of model design. The hybrid model combines multiple models, which helped capture different patterns in the data. This results in more stable and reliable predictions, especially when working with complex physiological signals. In addition, preprocessing steps such as filtering, smoothing, and subject-level normalization contribute to improving data quality and reducing differences between individuals. The subject-level normalization used in this work helps address one of the main challenges encountered, which is the variation between individuals. It was observed that different subjects have different physiological baselines, such as how their bodies responded to stress or how their skin reacted, which can affect model performance. By normalizing each subject's data based on their own rest state, the model becomes more adapted to each individual and less influenced by these differences. This makes the approach more reliable and more suitable for real-world applications, especially in personalized health monitoring systems that need to adapt to different users. This work highlights that carefully designed classical ML approaches could outperform more complex DL models, especially when working with relatively small datasets where DL models lack sufficient training samples to learn generalizable representations. This is an important insight, as it suggests that model simplicity and proper data handling could be more effective than increasing model complexity.

## VI. CONCLUSION

Stress affects an individual's physical and psychological well-being, impacting their quality of life. Long-term exposure to stress has been related to a wide range of health problems. As individuals are increasingly exposed to elevated levels of stress in both personal and professional contexts. The need for an effective stress-monitoring model is more critical than ever. Early detection of stress enables timely intervention, reduces the risk of long-term health complications, and supports overall well-being. However, the traditional approaches for stress assessment, such as self-reported questionnaires, are often subjective and may not accurately reflect an individual's true physiological state. This limitation drives the growing interest in developing objective, data-driven approaches for stress detection using physiological signals. This research work proposed a hybrid model that detects stress using EDA signals from the UBFC-Phys dataset. The EDA signals were preprocessed using low-pass filtering and smoothing, followed by subject-level normalization to reduce inter-individual variability. A total of 23 features were extracted, and a hybrid weighted soft-voting ensemble combining GB, RF, KNN, LR, and DT was used to classify stress into three levels. The proposed model achieved an AC of 92.16% in the multiclass classification task, with balanced PR, recall, and F1 values of 92.16%. In the binary classification task, a perfect AC of 100% was achieved. This approach provided a practical and efficient solution for real-world stress monitoring and continuous health-monitoring applications. Future work may concentrate on enhancing and extending the proposed approach by assessing the pipeline on larger and more diverse datasets, and applying cross-dataset validation to offer stronger evidence of its generalizability. Incorporating additional physiological signals such as HR and BVP alongside EDA, within a multimodal fusion framework, may further enhance classification accuracy, particularly for distinguishing between closely related stress states. Investigating advanced DL architectures with transfer learning strategies may also help address the limitations observed with DL models in this study.

## FUNDING

This scientific research study was derived from a research grant funded by Taibah University, Madinah, Kingdom of Saudi Arabia, with the grant number (1447-06).

## DATA AVAILABILITY

The data presented in this work are openly available in the UBFC-Phys repository and can be accessed via [https://ieee-dataport.org/open-access/ubfc-phys-2].

## ACKNOWLEDGMENT

The authors would like to acknowledge the support of Taibah University and the contributors of the UBFC-Phys dataset.

## REFERENCES


[1] Harvard Health Publishing, "Identifying and relieving stress," Harvard Health. Accessed: Nov. 16, 2025. [Online]. Available: https://www.health.harvard.edu/topics/stress

[2] American Psychological Association, "Stress effects on the body," APA. Accessed: Nov. 16, 2025. [Online]. Available: https://www.apa.org/topics/stress/body

[3] H. M. Morales-Fajardo et al., "Towards a Non-Contact Method for Identifying Stress Using Remote Photoplethysmography in Academic Environments," Sensors, vol. 22, no. 10, p. 3780, May 2022, doi: 10.3390/s22103780.

[4] J. C. Fruehwirth, M. E. Mazzolenis, M. A. Pepper, and K. M. Perreira, "Perceived stress, mental health symptoms, and deleterious behaviors during the transition to college," PLoS ONE, vol. 18, no. 6, p. e0287735, Jun. 2023, doi: 10.1371/journal.pone.0287735.

[5] H. Kerbage, O. Bazzi, W. El Hage, E. Corruble, and D. Purper-Ouakil, "Early interventions to prevent post-traumatic stress disorder in youth after exposure to a potentially traumatic event: A scoping review," in Healthcare, MDPI, 2022, p. 818. Accessed: Nov. 19, 2025. [Online]. Available: https://www.mdpi.com/2227-9032/10/5/818

[6] L. K. McCorry, "Physiology of the Autonomic Nervous System," Am J Pharm Educ, vol. 71, no. 4, p. 78, Sep. 2007, doi: 10.5688/aj710478.

[7] F. Rodrigues and H. Correia, "Semi-supervised and ensemble learning to predict work-related stress," J Intell Inf Syst, vol. 62, no. 1, pp. 77–90, Feb. 2024, doi: 10.1007/s10844-023-00806-z.

[8] R. M. Sabour, Y. Benezeth, P. De Oliveira, J. Chappé, and F. Yang, "UBFC-Phys: A Multimodal Database For Psychophysiological Studies of Social Stress," IEEE Transactions on Affective Computing, vol. 14, no. 1, pp. 622–636, Jan. 2023, doi: 10.1109/TAFFC.2021.3056960.

[9] S. Koldijk, M. Sappelli, S. Verberne, M. A. Neerincx, and W. Kraaij, "The SWELL Knowledge Work Dataset for Stress and User Modeling Research," in Proceedings of the 16th International Conference on Multimodal Interaction, Istanbul Turkey: ACM, Nov. 2014, pp. 291–298. doi: 10.1145/2663204.2663257.

[10] S. Chakraborty et al., "Deep Learning Based Stress Assessment Using PPG Signals from WESAD Dataset," in 2025 23rd IEEE Interregional NEWCAS Conference (NEWCAS), Jun. 2025, pp. 296–300. doi: 10.1109/NewCAS64648.2025.11106968.

[11] J. A. Healey and R. W. Picard, "Stress Recognition in Automobile Drivers." physionet.org, 2008. doi: 10.13026/C2SG6B.

[12] S. Hosseini et al., "A multimodal sensor dataset for continuous stress detection of nurses in a hospital," Sci Data, vol. 9, no. 1, p. 255, Jun. 2022, doi: 10.1038/s41597-022-01361-y.

[13] K. Mundnich et al., "TILES-2018, a longitudinal physiologic and behavioral data set of hospital workers," Sci Data, vol. 7, no. 1, p. 354, Oct. 2020, doi: 10.1038/s41597-020-00655-3.

[14] W.-K. Beh, Y.-H. Wu, An-Yeu, and Wu, "MAUS: A Dataset for Mental Workload Assessmenton N-back Task Using Wearable Sensor," Nov. 03, 2021, arXiv: arXiv:2111.02561. doi: 10.48550/arXiv.2111.02561.

[15] G. Vos, K. Trinh, Z. Sarnyai, and M. Rahimi Azghadi, "Ensemble machine learning model trained on a new synthesized dataset generalizes well for stress prediction using wearable devices," Journal of Biomedical Informatics, vol. 148, p. 104556, Dec. 2023, doi: 10.1016/j.jbi.2023.104556.

[16] Y. Hasanpoor, K. Motaman, B. Tarvirdizadeh, K. Alipour, and M. Ghamari, "Stress detection using ppg signal and combined deep cnn-mlp network," in 2022 29th National and 7th International Iranian Conference on Biomedical Engineering (ICBME), IEEE, 2022, pp. 223–228. Accessed: Nov. 19, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10052957/

[17] X. Zhang et al., "Dynamic Alignment and Fusion of Multimodal Physiological Patterns for Stress Recognition," IEEE Transactions on Affective Computing, vol. 15, no. 2, pp. 685–696, Apr. 2024, doi: 10.1109/TAFFC.2023.3290177.

[18] A. Arsalan, M. Majid, S. M. Anwar, and U. Bagci, "Classification of perceived human stress using physiological signals," in 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC), IEEE, 2019, pp. 1247–1250. Accessed: Nov. 19, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8856377/

[19] MD. N. Al Hasan, N. S. Noman, M. H. Chowdhury, and A. Chowdhury, "Improving Multimodal Stress Detection Using MAUS Dataset with PPG, ECG and GSR Signal," in 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE), Feb. 2025, pp. 1–5. doi: 10.1109/ECCE64574.2025.11013834.

[20] A. Mathur, Shikha, and D. Sethia, "Body Sensor-Based Multimodal Nurse Stress Detection Using Machine Learning," in 2024 16th International Conference on COMmunication Systems & NETworkS (COMSNETS), Jan. 2024, pp. 67–73. doi: 10.1109/COMSNETS59351.2024.10427006.

[21] S. P. Kar, N. K. Rout, and J. Joshi, "Assessment of Mental Stress From Limited Features Based on GRU-RNN," in 2021 IEEE 2nd International Conference on Applied Electromagnetics, Signal Processing, & Communication (AESPC), IEEE, 2021, pp. 1–4. Accessed: Nov. 19, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9708506/

[22] M. Gil-Martin, R. San-Segundo, A. Mateos, and J. Ferreiros-Lopez, "Human stress detection with wearable sensors using convolutional neural networks," IEEE Aerospace and Electronic Systems Magazine, vol. 37, no. 1, pp. 60–70, 2022.

[23] Z. Liu, Z. Shi, G. Zhang, J. Jung, and M. Li, "Mental Stress Detection Using PPG Signals Based on Transformer-LSTM Model," in 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Dec. 2024, pp. 7095–7097. doi: 10.1109/BIBM62325.2024.10822335.

[24] S. Yang et al., "A deep learning approach to stress recognition through multimodal physiological signal image transformation," Sci Rep, vol. 15, p. 22258, Jul. 2025, doi: 10.1038/s41598-025-01228-3.

[25] G. Gupta, M. Khan, K. I. Sherwani, and Manaullah, "Overview of the effect of physiological stress on different biological signals," in 2023 International Conference on Recent Advances in Electrical, Electronics & Digital Healthcare Technologies (REEDCON), May 2023, pp. 251–255. doi: 10.1109/REEDCON57544.2023.10150828.

[26] D. Makowski et al., "NeuroKit2: A Python toolbox for neurophysiological signal processing," Behav Res, vol. 53, no. 4, pp. 1689–1696, Aug. 2021, doi: 10.3758/s13428-020-01516-y.

[27] W. H. Alawee, A. Basem, and L. A. Al-Haddad, "Advancing biomedical engineering: Leveraging Hjorth features for electroencephalography signal analysis," Journal of Electrical Bioimpedance, vol. 14, no. 1, pp. 66–72, Jan. 2023, doi: 10.2478/joeb-2023-0009.

[28] R. Valerio and M. Mahmoud, "A multimodal Framework for exploring behavioural cues for automatic Stress Detection," in Proceedings of the 27th International Conference on Multimodal Interaction, Canberra Australia: ACM, Oct. 2025, pp. 535–539. doi: 10.1145/3716553.3750797.