IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Optimization of Wearable Biosensor Data for Stress Classification Using Machine Learning and Explainable AI

SHIKHA[1], DR. DIVYASHIKHA SETHIA[2], and Prof. S. Indu[3]
[1]Department of Computer Science Engineering, Delhi Technological University, India (e-mail: shikha.9126@gmail.com)
[2]Department of Software Engineering, Delhi Technological University, India (e-mail: divyashikha@dtu.ac.in)
[3]Department of Electronics and Communication, Delhi Technological University, India (e-mail: s.indu@dce.ac.in)

Corresponding author: Prof. S. Indu (s.indu@dce.ac.in).

This paragraph of the first footnote will contain support information, including sponsor and financial support acknowledgment. For example, "This work was supported in part by the U.S. Department of Commerce under Grant BS123456."

ABSTRACT This work utilizes wearable devices for real-time stress detection and investigates the effectiveness of meditation audio in reducing stress levels after academic exposure. Physiological data, including Interbeat Interval (IBI)-derived Heart Rate Variability (HRV), Blood Volume Pulse (BVP), and electrodermal activity (EDA), are collected during the Montreal Imaging Stress Task (MIST). The stress classification methodology employs an integrated approach using Genetic Algorithm and Mutual Information to reduce feature set redundancy. It further uses Bayesian optimization to fine-tune machine learning hyperparameters. The results indicate that the combination of EDA, BVP, and HRV achieves the highest classification accuracy of 98.28% and 97.02% using the Gradient Boosting (GB) algorithm for 2-level and 3-level stress classification. In contrast, EDA and HRV alone achieve a comparable accuracy of 97.07% and 95.23% for 2-level and 3-level stress classification, respectively. Furthermore, the SHAP Explainable AI (XAI) analysis confirms that HRV and EDA are the most significant features for stress classification. The study also finds evidence that listening to meditation audio reduces stress levels. These findings highlight the potential of wearable technology combined with machine learning for real-time stress monitoring and management in academic environments.

INDEX TERMS Academic environment, Explainable AI, Feature Selection, Machine Learning, Mental stress, Wearable Device

I. INTRODUCTION

Students frequently face stressful situations in today's demanding academic environment, from class presentations to academic examinations. According to Prabu et al., [1], "academic stress is the psychological and emotional strain that arises from the demands and expectations associated with formal education and scholarly pursuits". This combination of rigorous academic workloads and competitive social dynamics can lead to significant mental stress, affecting their overall well-being and academic performance [2]. Studies have shown that unmanaged academic stress can lead to decreased motivation, anxiety, sleep problems, and even physical health issues [3] [4] [5]. Recognizing and addressing students' challenges in managing academic stress is crucial, as this has become a critical aspect of their educational journey.

However, the current focus in healthcare often lies in addressing and treating existing mental health issues and their associated symptoms. This approach is vital, but it is equally important to acknowledge the subtle indicators of stress and anxiety before they escalate. Recognizing these early signs empowers students to take action and manage their stress effectively, potentially preventing the need for more intensive interventions later.

To effectively identify these early signs of stress, researchers are increasingly emphasizing the relationship between variations in physiological signals and stress. When individuals undergo these physical and mental reactions to stress, their physiological signals also change [6] [7]. For example, Holtz et al. [8] introduced a sudden air horn sound and discovered that stress responses, such as an increase in Respiration Rate (RESP), Heart Rate (HR), and Electrodermal Activity (EDA) signals, occurred. Hence, it is vital to understand the connections between variations in physiological signals and stress to identify stress effectively. Concur-

rently, there is a growing emphasis on using scientific and technological methods to monitor and analyze physiological signals, making it a significant research area in academia and industrial settings.

Wearable technology is promising for continuously and remotely monitoring mental health [9] [10]. These devices can gather physiological patient data and offer valuable contextual information. Additionally, the power of machine learning has enhanced data processing speed and the ability to derive meaningful insights from this physiological data. Recent advancements in machine learning have introduced ensemble techniques such as DeepAVP-TPPred [11], iAFPs-Mv-BiTCN [12], Deepstacked-AVPs [13], AIPs-SnTCN [14], and pAtbP-EnC [15], which have shown promising results in various applications. These models precisely predict peptide functionality by extracting features from sequential and evolutionary descriptors through ensemble learning techniques. These advanced predictors offer robust frameworks for analyzing physiological signals, providing more accurate and early detection of stress and anxiety indicators. One notable category of wearable devices is smartwatches. These devices function like miniaturized smartphones, boasting impressive computational capabilities. Moreover, they come equipped with various sensors capable of collecting physiological signals such as EDA, Photoplethysmography (PPG), Electrocardiography (ECG), and Skin Temperature (ST). This combination of technology and sensors makes them a valuable tool for monitoring mental health remotely and continuously.

Moreover, researchers have employed various stress-inducing tasks, including public speaking [16], questionnaires, programming contests [17], ice tasks, and games [18] while recording physiological signals. Some researchers have also delved into stress classification using logical thinking-based tasks such as pursuing mental arithmetic and the Montreal Imaging Stress Task (MIST) test [19].

The MIST, a well-validated and progressive stress-inducing test, has been extensively used for stress classification. It involves three distinct stages: resting, control, and experimental. The participants solve a series of mathematical tasks without time constraints during the control stage. However, in the experimental stage, participants are presented with challenging arithmetic problems that are time-constrained and aimed at inducing a higher stress level.

Various researchers [20] [21] explored EEG-based stress classification experiments using MIST. Minguillon et al. [22] proposed a portable system for real-time detection of stress levels. They utilized Electroencephalography (EEG), ECG, Electromyography (EMG), and Galvanic Skin Response (GSR) signals recorded during the MIST task. Using EEG alone, they achieved only 50% accuracy, but when combined with multimodal data, the accuracy increased to 86%. The results show the efficacy of multimodality, as it provides a comprehensive understanding of stress levels. Although EEG offers valuable insights into brain activity, it has some practical limitations, including noise susceptibility and discomfort, which hinder its everyday use. Wearable devices, especially smartwatches, emerge as a promising alternative for continuous, non-invasive monitoring of physiological signals like heart rate variability (HRV) and EDA. The unobtrusive nature of the signals makes them well-suited for real-time stress assessment in academic environments.

Table 1 shows the effectiveness of stress classification studies during tasks like MIST and incorporates wearable devices to record physiological signals such as EEG, ECG, EDA, and RESP. Notably, numerous studies explored EDA signals for stress classification.

Setz et al. [24] employed EDA and MIST as a stressor. The authors investigated six classifiers, including SVM (linear, Quad, RBF, Polynomial), LDA (linear discriminant analysis), and KNN (K-nearest Neighbours). They concluded that monitoring EDA allows discrimination between cognitive load and stress with an accuracy larger than 80% with leave-one-person-out cross-validation.

In another study, Zhu et al. [27] explored stress classification using EDA signal and MIST. They obtained the highest accuracy of 92.9% using the SVM classifier. Furthermore, Hsieh et al. [28] proposed a feature selection framework based on EDA signal. They conclude that the EDA signal contains the most significant information for stress detection and obtained the highest accuracy of 92.38%.

Despite the extensive use of EDA in stress classification studies, there needs to be more literature regarding the combined use of ECG and EDA in conjunction with MIST. Han et al. [25] explored the combination of ECG and RESP with MIST for stress classification and obtained the highest accuracy of 96% using RF. However, RESP may be less reliable as a stress biomarker due to challenges in accurately measuring breath rate in real-life environments and its susceptibility to confounding factors beyond stress, such as physical exertion, respiratory illnesses, changes in posture, or environmental conditions [35].

To the best of our knowledge, in the context of academic learning, no research has specifically investigated the combination of ECG and EDA during MIST. ECG provides insight into autonomic nervous system activity through HRV features, while GSR reflects sympathetic nervous system arousal through skin conductance changes [36]. Stress activates the sympathetic nervous system, leading to increased EDA and decreased HRV. Therefore, changes in these signals can help detect stress when an individual encounters a stressor. EDA is more sensitive when capturing acute stress [37] [38]. In contrast, HRV provides a more stable measure of chronic stress [39]. Combining data from EDA and IBI-derived HRV enables the development of more precise and resilient stress detection systems.

A. CONTRIBUTIONS

The study investigates the feasibility of IBI-derived HRV and EDA data in an academic setting by collecting student physiological data during MIST and analyzing the effectiveness of listening to meditation audio in alleviating stress. The study also employs an integrated feature selection, Genetic

TABLE 1. Previous work on MIST and MAT stressors using Physiological Signals

Author	Signal	No. of Sub	Stressor Type	Phases	Experiment Duration	Classifiers	Result	Limitations
Minguillon et al. [22]	EEG, EMG, GSR, ECG	10	MIST	Relax, stress, Neutral	30 minutes	LDA	EEG Acc: 50%, Combined Sensors Acc (EEG, EMG, ECG, GSR): 86%	- only 4 channels are used. Did not work on Time Domain and Time-Frequency Domain features
Xia et al. [23]	EEG, ECG	22	MIST	Rest–Control-Experimental-Recovery	80 minutes	SVM	Acc: 79.54%, Sen: 81.00%, Spec: 78.00%	- Experiment duration is too long. Did not work on Time Domain and Time-Frequency Domain features
Setz et al. [24]	EDA	33	MIST	Rest, Control, Experiment, Recovery	50 minutes	SVM (Quad, RBF, Polynomial), LDA, Nearest class center	Accuracy: 82.8%	- Experiment Duration is too long. Utilized only EDA
Han et al. [25]	ECG, RSP	39	MIST	Rest - Control - Experiment - Recovery	18 minutes (excluding stress report filling time)	SVM, RF, LDA, Adaboost, KNN	3-level Acc:84% 2-level Acc:96%	- Can Utilize other physiological signals from wearable sensors.
Gjoreski et al. [26]	IBI, RSP, BVP, ST, EDA	22	MAT	Easy - Medium - Tough	30 minutes	SVM, RF, Boosting, Bagging, KNN, Naive Bayes, Decision Tree, Ensemble Selection	Recall:70% Precision:95%	- Considered only male participants.
Zhu et al. [27]	EDA	112	MAT	Easy - Medium - Tough	—	KNN, SVM, NB, LR, RF	Accuracy: 92.9%	- Other physiological signals results could be compared - Other evaluation metrics could be utilized
Hsieh et al. [28]	EDA	15	Public Speaking and MAT	Stress - Non stress	2 hr	XGboost	Accuracy: 92.38%	- Less sample size - Experimental duration is too long
Siirtola et al. [29]	ST, BVP, HR	15	Public Speaking and MAT	Stress - Non-stress	2 hr	LDA, QDA, RF	Accuracy: 87.4%	- Accuracy can be improved - Less sample size - Experimental duration is too long
Kumar et al. [16]	BVP, ECG, EDA, EMG	15	Public Speaking and MAT	Stress - Non-stress	2 hr	CNN hierarchal model	Accuracy: 87.77% F1 score: 83%	- Accuracy can be improved - Less sample size - Experimental duration is too long
Jaiswal et al. [30]	GSR	42	TSST, MAT,	Stress - Non stress	25 minutes	RF, SVM, KNN and AdaBoost.	Accuracy: 80%	- Other physiological biomarkers can be utilized.
Androutsou et al. [31]	PPG, GSR	32	SCWT, MAT	Rest, Control, Stress	45–60 min	SVM, KNN, DT, RF, XGB, RUSB, LGBM, ADB	LGBM Accuracy: 90.06% F1score: 90%	- reliance on self-reported stress levels, which may introduce subjective bias into the results
Campanella et al. [32]	PPG, EDA	29	TSST, MIST, Puzzle solving	Rest, Stress	30-35 min	RF, SVM, Log Regression	RF: 76.5%	- No hyperparameter tuning is adopted. - Imbalanced dataset.
Hemakom et al. [33]	ECG, EEG	66	MIST	Rest, Control, Stress	60 min	KNN, RBF-SVM, RF, Adaboost, MLP	87.60%	- Different models are trained for different stress levels which may result in misclassification in real-world scenarios.
Benchekroun et al. [34]	HRV	74 + 27	SCWT, MAT	Rest, Stress, Recovery	55 min	RF, LR	81%	- Self-report data label in second dataset could lead to biases. - Other physiological signals could be used.

Algorithm, and Mutual Information (GA+MI) to select the most important features from the data. Further, Bayesian optimization is used to tune machine learning classifiers and predict the stress level of each individual [40] [41]. The results obtained are validated using ten-fold cross-validation techniques. Lastly, the system's performance is analyzed for each individual to validate the impact of meditation audios on stress alleviation.

1) To conduct an innovative study to collect physiological signals such as EDA, IBI, and BVP from wearable devices across different stress levels and meditation phases.
2) To develop and optimize machine learning models for stress detection by implementing a hybrid feature selection approach combined with Bayesian optimization.
3) To employ the SHAP Explainable AI (XAI) technique to identify the most influential physiological features significantly impacting stress detection, promoting transparency in machine learning predictions.
4) To illustrate the impact of meditation on stress reduction through dynamic visualizations of physiological data collected from wearable devices among college students.

B. ROADMAP

The structure of this paper is as follows. Section 2 describes the experimental methodology employed in the study, which includes a detailed description of the data acquisition, preprocessing, and the proposed protocol design for developing a stress detection model. Section 3 presents the experimental results, and section 4 comprehensively discusses the results. Lastly, Section 5 explains the concluding remarks on this paper and potential future research directions.

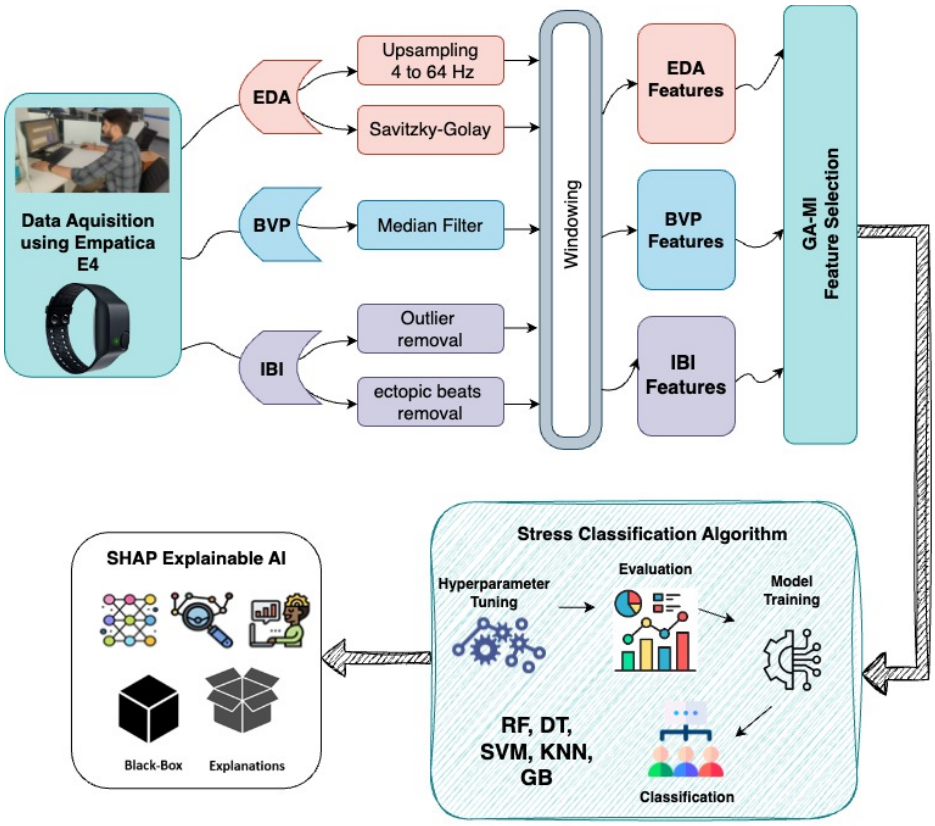

FIGURE 1. Proposed Classification Methodology

II. EXPERIMENTAL METHODOLOGY

This paper structures this section into five subsections shown in Figure 1. The methodology involves data acquisition, which outlines the procedure for collecting physiological signals, followed by data preprocessing to enhance signal quality. Subsequently, Feature Extraction and Feature Selection are performed. Further, the selected features undergo statistical analysis, and then these features are input into distinct classifiers, which are optimized using Bayesian optimization.

A. DATA AQUISITION:

This section involves the participant's details and the procedure during stress data collection.

1) Participants:

The study collects data from 36 engineering students aged between 17 and 25 (thirty-two males and four females) at Delhi Technological University, India. However, due to some participants missing IBI data caused by movements, only 30 student's information is considered for the study. During the data collection procedure, the participants are briefed about an overview of the study protocol and its objectives and asked to sign a consent form before starting the data collection.

The study adheres to strict ethical guidelines to ensure the well-being of participants. The DTU ethics committee, Delhi Technological University, Delhi, approves these guidelines. Before participation, all students are informed of the potential stress-inducing nature of the experiment, provide informed consent, and are asked to complete the Perceived Stress Scale (PSS) questionnaire to assess their self-reported stress levels [42]. Participants can also withdraw from the study at any time without penalty. Additionally, trained personnel monitored participants throughout the experiment and conducted debriefing sessions to address any concerns or discomfort experienced during the procedure. The test comprises five phases, as shown in Figure 2: Neutral phase, Rest phase, Controlled phase, Experimental phase, and Recovery phase.

1) **Neutral Phase:** The Neutral phase serves as the training phase, introducing participants to the MIST test and allowing them to familiarize themselves with the test's controls and answer format. Questions in this test have single-digit answers ranging from 0 to 9, and participants respond using a rotatory dial on the screen, controlled by the mouse's left and right buttons.
2) **Rest Phase:** During the rest phase, participants are instructed to sit comfortably in a relaxed position with minimal movements while a blank screen is displayed. This phase aims to establish a baseline physiological state for each participant.
3) **Controlled Phase:** In this phase, participants are presented with arithmetic questions of differing difficulty levels and no time constraints. They are encouraged to

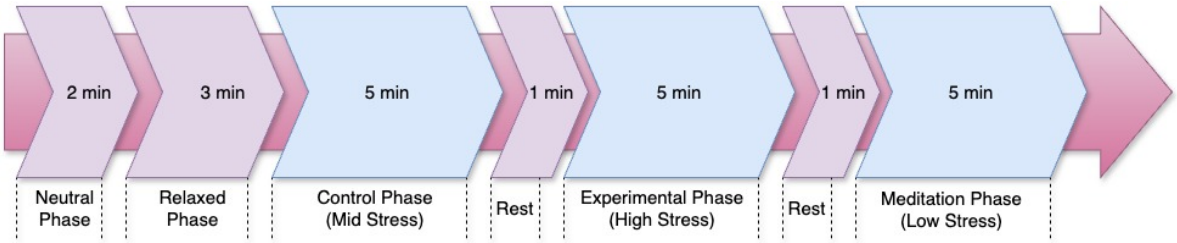

FIGURE 2. **Data Acquisition Procedure**

respond to as many questions as accurately as possible without receiving feedback. The study labels this phase as 'mild stress.'

4) **Experimental Phase:** The Experimental Phase presents adaptive arithmetic questions with an added element of social threat. Each question is timed, and participants are shown their current performance score alongside an artificially inflated average score of all participants. This artificial average score is intentionally set higher than the participant's achievable score, creating a sense of competition and heightened stress. Additionally, incorrect or untimely responses increase the time limit for the subsequent question, while three consecutive correct answers decrease the time limit for the next question. This phase effectively elicits greater stress in participants and is classified as 'high stress.'
5) **Recovery Phase:** The final phase assesses physiological changes as participants return to a relaxed state. Participants are engaged in a stress alleviation task involving listening to guided meditation audio with closed eyes while sitting comfortably. The study labels this phase as 'low stress'.

The current study includes a 2-minute training phase, followed by a 5-minute control section and a 5-minute experimental section. One-minute rest sections are interspersed between the control and experimental phases. The study concludes with a 5-minute recovery section for stress alleviation.

2) Empatica E4

This study continuously monitors physiological signals throughout the experiment using the Empatica E4 wristwatch. It consists of four sensors: a temperature sensor, an accelerometer, EDA sensors, and PPG sensors. Further, the PPG Sensor provides IBI and BVP signals. The collected data is transmitted via Bluetooth to a connected smartphone and monitored in real-time through the E4 Real-time app. Subsequently, this data is uploaded to Empatica's cloud platform, E4 Connect, for initial processing. The E4 device generates .csv files containing a variety of physiological measures, including skin temperature (ST), blood volume pressure (BVP), inter-beat-interval (IBI), EDA, heart rate (HR), and acceleration. Notably, the sampling rate for ST and EDA is 4 Hz, while the sampling rate for IBI, HR, and BVP is 64 Hz.

B. DATA PREPROCESSING

The primary objective of the current work is to explore physiological signals using the Empatica E4 and evaluate the accuracy of the proposed model using machine learning techniques. Even though the study collects the data inside the laboratory, its quality may be susceptible to various factors, such as sensor detachment, outliers, and missing information. Consequently, signal preprocessing is deemed necessary to improve the signal-to-noise ratio and safeguard valuable information from inadvertent filtering.

Labeling: Before the data preprocessing step, the data is clipped based on consecutive sessions. In three-level stress classification, the study labels the control phase as 'Mid Stress,' the experimental phase is labelled 'High Stress,' and the recovery phase as 'Low Stress.'

Signal Windowing: Signal windowing is a powerful technique to extract more detailed information from time-series data. This study implements a ten-second rolling window with a 50% overlap on the wristband data, maximizing the information extracted from the continuous data stream.

1. IBI (Interbeat Interval): The IBI.csv file consists of time intervals between successive heartbeats, and it is derived from the PPG Sensor. The data is preprocessed and analyzed using the *hrvanalysis* module in Python, which offers a comprehensive set of Heart Rate Variability (HRV) analysis tools [43]. These tools encompass preprocessing, R-peak detection, and feature extraction.

Data Cleaning: The IBI data is initially examined for missing values or outliers. These outliers are handled by employing interpolation techniques. Further, the IBI series may exhibit gaps due to discarded ectopic beats or undetected heartbeats, a common occurrence with wearable devices that can impact HRV analysis accuracy. These ectopic beats are addressed using the *"malik"* method [44].

2. EDA (Electrodermal Activity): As the five physiological signals are recorded at different sampling frequencies, it is essential to ensure that all the signals have the same sample size before analyzing [32]. The study employs upsampling, which increases the number of signal samples to address this sampling frequency discrepancy. Notably, among all the recorded features, the BVP and IBI are sampled at the highest frequency of 64 Hz. Consequently, this study upsamples the EDA signal from 4 to 64 Hz to standardize the sampling frequency for both signals. In addition, to eliminate potential

TABLE 2. 58 Multi-domain features derived from EDA, IBI, and BVP signal

Feature Domain	Feature Description	Features	Feature ID
GSR-Statistical	Captures central tendencies and variability, providing important information about overall arousal levels and autonomic activity.	var_gsr, mean_gsr, kurtosis_gsr, skew_gsr, std_gsr	1-5
GSR-Tonic	Represents slowly varying baseline characteristics of the GSR signal, capturing continuous variations in skin conductance levels and reflecting an individual's overall arousal state and tonic sympathetic activity.	SCL_slope (tonic), var_scl, mean_scl, kurtosis_scl, skew_scl, slope_scl, std_scl	6-12
GSR-Phasic	Captures rapid and short-term fluctuations in GSR signal represent momentary responses to stimuli or events and reflect the phasic sympathetic activity.	scr_peaks (phasic), var_scr, mean_scr, kurtosis_scr, skew_scr, max_scr, std_scr	13-19
HRV-Time Domain	Statistical measures derived from the time intervals between consecutive heartbeats (RR intervals) in an HRV analysis.	median_nni, mean_nni, pnni_50, nni_50, nni_20, pnni_20, hrv_sdnn, hrv_rmssd, hrv_sdsd, range_nni, cvnni, cvsd, max_hr, mean_hr, std_hr, min_hr, tinn, triangular_index, cvi, sd1, csi, sd2, sampen, ratio_sd1_sd2, HR_max-HR_min	20-44
HRV-Frequency Domain	Captures the power distribution across different frequency bands, representing the relative contributions of sympathetic and parasympathetic activities in heart rate regulation.	total_power, lf, vlf, hf, lf_hf_ratio, hfnu, lfnu	45-51
BVP-Statistical Features	Numerical values that describe the BVP signal and its changes over time and frequency	Mean BVP, standard deviation, maximum, minimum, power VLF BVP, power LF BVP, power HF BVP	52-58

artefacts, the study applies a Savitzky-Golay [45] filter with a 40-point window and a sigma of 400 ms to smoothen the signal [46]. Subsequently, all the artefact-free segments underwent the feature extraction process.

3. BVP (Blood Volume Pressure):

The study derives the Blood Volume Pulse (BVP) signal from the PPG sensor by passing it through a high-pass filter. The specific cut-off frequency for this filter can vary, but it is typically set between 0.05 and 0.5 Hz. The data in this file represent BVP values calculated by a built-in algorithm, measured in nanowatts (nWatt) units. The BVP signals undergo filtering using a median filter, which involves scanning through the signal, point by point, and replacing each value with the median of neighbouring entries.

C. FEATURE EXTRACTION

The study retrieves numerical features from each data window in the feature extraction phase. Features are obtained by statistical functions and time and frequency analysis methods depending on the specific signal type.

The Empatica E4 device supplies IBI data, which is then employed to extract HRV features. The study gathers HRV features across various domains, including time, frequency, and non-linear, using the *hrvanalysis* Python module. In addition, statistical features are extracted from the BVP data using *numpy* module. The study employs the *PyEDA* package in the Python environment to extract phasic and tonic features from the processed EDA signal. Table 2 presents a total of 58 extracted features, encompassing HRV, EDA, and BVP, categorized by their respective domains. Additionally, Table 3 shows sample data example for 2-level stress classification.

D. FEATURE SELECTION

Feature selection is crucial in data analysis, particularly in clinical diagnosis systems [47]. Its primary goal is to identify and retain the most relevant features while discarding those that contribute minimally to the classification task. This process simplifies the data input, enhances model performance, and provides insights into the key factors influencing accurate stress classification. The primary advantage of the Feature Selection Algorithm (FSA) is reduced data complexity, improved model performance, enhanced interpretability, and reduced computational cost [48].

The success of feature selection relies on two critical factors: the search strategy and the selection criteria. Various methods are employed in different feature selection approaches to create subsets and drive the search processes. Extensive comparative studies have revealed that, especially in the case of large finite spaces, the search for the optimal solution becomes computationally demanding due to the exponential search space it entails. Consequently, most search strategies aim to identify sub-optimal or nearly optimal solutions. Recent research interest has gravitated towards global search algorithms, known as metaheuristic techniques.

The feature selection process's efficacy depends significantly on considering two key aspects: the search strategy and the criteria employed [49]. Various approaches to feature selection use distinct methods to generate subsets and advance the search processes. Comparative studies [50], [51], [52] have revealed that, particularly in large finite spaces,

TABLE 3. Sample Data Example for 2-Level Stress Classification with Selected Features

GSR Statistical	GSR-Tonic		GSR Phasic	HRV Time domain			HRV Frequency domain	BVP Statistical	
mean_gsr	mean_scl	slope_scl	std_scr	RMSSD RR	Mean HR	Min HR	LF/HF	Mean BVP	Label
0.037	0.039	0.017	0.002	0.190	0.517	0.511	0.064	0.423	0
0.039	0.041	0.022	0.002	0.002	0.528	0.532	0.996	0.438	0
0.040	0.039	0.001	0.024	0.002	0.531	0.534	0.987	0.442	0
0.035	0.037	0.002	0.005	0.118	0.129	0.125	0.300	0.428	1
0.036	0.037	0.001	0.003	0.118	0.125	0.124	0.181	0.415	1
0.033	0.035	0.006	0.000	0.052	0.484	0.485	0.063	0.411	0
0.038	0.040	0.014	0.005	0.094	0.098	0.098	0.438	0.381	1
0.040	0.041	0.028	0.006	0.045	0.112	0.111	0.500	0.434	1
0.037	0.038	0.000	0.006	0.013	0.095	0.095	0.993	0.446	1
0.037	0.038	0.004	0.007	0.013	0.096	0.097	0.994	0.455	1
0.035	0.037	0.008	0.003	0.000	0.506	0.509	0.991	0.412	0

seeking the optimal solution is computationally demanding due to the resulting exponential search space. Consequently, most search strategies aim to identify sub-optimal or nearly optimal solutions [53]. Recent research has increasingly focused on global search algorithms, also known as metaheuristic algorithms. These strategies aim to find an optimal solution by exploring the entire space rather than limited feature spaces. Metaheuristic algorithms prove especially effective when dealing with uncertain and dynamic information. Hence, these advanced approaches efficiently manage the uncertainty and imprecision inherent in clinical data and medical images, particularly in biological predictive systems. Genetic Algorithm (GA) emerges as a natural choice for addressing feature selection challenges, and it is one of the most widely utilized optimization methods for navigating complex and non-linear search spaces [47].

1) Genetic Algorithm (GA)

GA represents a prominent branch of evolutionary computing and is a rapidly expanding field within Artificial Intelligence (AI). It is worth noting that the evolutionary computation community does not universally approve of a precise definition of "Genetic Algorithms" [54]. Nevertheless, GAs can be described as adaptive heuristic search algorithms, originally conceived by Holland in the 1960s, drawing inspiration from Darwin's theory of evolution, specifically, the concept of "survival of the fittest" [55]. Five fundamental factors can influence the execution of GAs: population creation, fitness function, selection schema, genetic operators, and stopping criteria. The algorithm commences with an initial population of binary strings known as chromosomes, potential candidates for solving optimization problems. In each iteration, these populations undergo evaluation based on their fitness, followed by applying crossover and mutation genetic operators to favour more robust solutions. Typically, candidate solutions are encoded within binary strings composed of 0s and 1s. In this representation, 0 denotes that a particular feature has not been selected, while 1 signifies the inclusion of the corresponding feature.

A study by Ramdania et al. [56] presents compelling evi-

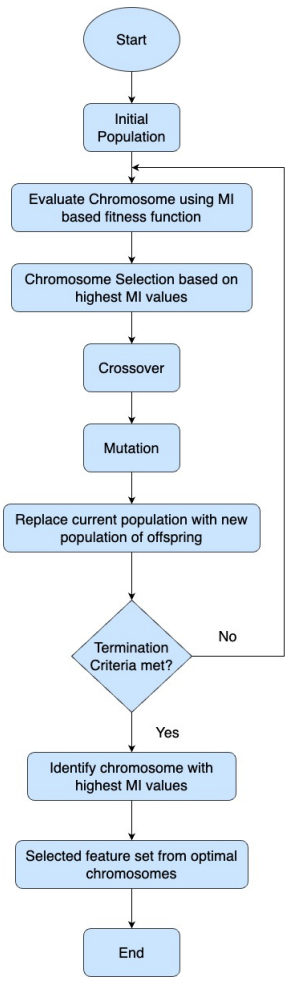

FIGURE 3. GA-MI Feature Selection Algorithm

dence that Genetic Algorithms excelled over alternative methods in efficiently identifying nearly optimal features within vast datasets. Building on this, Rostami et al. [57] introduced a hybrid algorithm designed to enhance solutions nearest to those identified by the Genetic Algorithm. In another study, a comparative analysis of algorithms for feature selection in pattern recognition was carried out [58]. The findings concluded that Genetic Algorithms excel particularly in handling large-scale problems.

2) Mutual Information (MI)

In research, there is a need to carefully consider the criteria used to evaluate classification performance to perform effective feature selection. Feature selection methods rely on evaluation criteria and search strategies to identify the optimal feature set. MI has gained significant attention in this context because it excels at measuring the correlation between features and class labels, and it is less sensitive to noise or outliers than other methods [59]. MI-based feature selection algorithms aim to select the best subset of features by maximizing the joint MI between input features and the target output. However, estimating high-dimensional MI is complex and computationally demanding, limiting its practical use [60]. As a result, many methods opt for low-dimensional MI. Over time, numerous MI-based feature selection algorithms have been developed in this field.

3) Genetic Algorithm with Mutual Information (GA-MI)

The study elevates the MI-based feature selection algorithm by integrating a global search algorithm for stress classification. Figure 3 shows the core operation of the Genetic Algorithm incorporating Mutual Information in the context of the physiology-based stress dataset. Further, Table 4 summarizes the parameters used for implementing the Genetic Algorithm FSA. The detailed procedure is outlined as follows:

1. *Create initial population:* Define a population size p, representing the number of chromosomes in the initial population. Each chromosome c_i is a binary string of length n, where n is the total number of features in the physiology-based stress dataset. Each bit c_{ij} in the string corresponds to a specific feature f_j. For each bit c_{ij} in the chromosome c_i, randomly set its value to 0 or 1. This random initialization ensures that the initial population contains diverse potential solutions.

2. *Compute Mutual Information:* MI quantifies the statistical dependence between two random variables. In this context, the second step measures the relevance of a set of features, represented by a chromosome c_i, to the target class y. A higher MI value indicates a stronger association between the features and the class. Given a chromosome c_i, let X_{c_i} denote the set of features selected by c_i (i.e., features corresponding to bits set to 1 in c_i. The mutual information MI (c_i,y) is calculated as follows:

$$MI(c_i, y) = \sum_{x \epsilon c_i} \sum_{y \epsilon Y} p(x, y) log \frac{p(x, y)}{p(x)p(y)} \tag{1}$$

where p(x,y) denotes the joint probability distribution of feature values $x \in X_{ci}$ and class labels y, p(x) represents the marginal probability distribution of feature values $x \in X_{ci}$, and p(y) represents the marginal probability distribution of class labels y. As a result, the highest MI value chromosomes are used in the next step.

3. *Fitness-Proportional Selection:* The selection probability $P(c_i)$ for each chromosome c_i is directly proportional to its mutual information $MI(c_i, y)$. It implies that chromosomes with higher MI values are more likely to be selected for reproduction, ensuring that the algorithm prioritizes features with greater relevance to the target class.

$$P(c_i) = \frac{MI(c_i, y)}{\sum_{j=1}^{p} MI(c_j, y)} \tag{2}$$

Where $MI(c_i, y)$ represents the mutual information of chromosome c_i, and the denominator represents the sum of mutual information values for all chromosomes in the population.

4. *Crossover and Mutation:* The crossover operator mimics genetic recombination by combining genetic information from two selected chromosomes (parents) to generate new offspring chromosomes. This exchange of information can potentially lead to the discovery of superior feature combinations. The mutation operator introduces random alterations to individual bits in the offspring chromosomes, potentially introducing new features or removing irrelevant ones. This mechanism helps the algorithm escape local optima and explore broader solutions.

5. *Iteration and Termination:* The algorithm iteratively executes steps 2 to 4, evolving the population over generations. The termination condition reaches the maximum number of generations G_{max}, typically determined experimentally based on the problem complexity and desired convergence.

6. *Optimal Solution:* The optimal feature c_{best} is identified as the chromosome with the highest mutual information $MI(c_{best}, y)$ in the final generation. It represents the features most relevant to the target class and is expected to yield the best classification performance.

E. ANALYSIS OF VARIANCE (ANOVA)

To further investigate the significance of the selected features, the study proceeded with an analysis using Analysis of Variance (ANOVA) test. ANOVA compares the differences in means among groups by examining the variance within and between groups, helping to determine if there are significant variations. This method assumes that observations are independent, normally distributed, and have equal variances among groups. It produces an F-statistic to assess significance of group differences. In this study, ANOVA is used to evaluate the distinctiveness of the selected feature subset by calculating degrees of freedom, p-values, and F-values for stress and non-stress classes. These statistical test parameters and the F-value are computed using following equations.

$$F = \frac{MS_B}{MS_w} \tag{3}$$

TABLE 4. Parameters for Implementing Genetic Algorithm

Parameter	Description	Value
Initial Population	Number of chromosomes in the initial population	100
Evaluation Criteria	Metric used to assess chromosome fitness	Mutual Information (MI)
Reproduction Operators	Genetic operators for generating new offspring	Single-point crossover, mutation rate: 0.05
Crossover Rate	Probability of crossover occurring between chromosomes	0.8
Selection Criteria	Method for selecting chromosomes based on fitness	Fitness-proportional selection with elitism
Termination Criteria	Condition for stopping the GA iterations	100 generations or convergence

$$df_B = k - 1 \tag{4}$$

$$df_w = N - 1 \tag{5}$$

$$MS_B = \frac{Sumofsquarebetweensample}{df_B} \tag{6}$$

$$MS_w = \frac{Sumofsquarewithinsample}{df_w} \tag{7}$$

The p-value and the F-value have an inverse relationship – when one is low, the other tends to be high. Specifically, when the p-value is less than 0.05, it implies that each feature contributes individual and valuable information to the results.

F. CLASSIFICATION ALGORITHM

After feature selection, the data is divided into a ratio of 80% for training and 20% for testing. Further, the study employs five classification algorithms: Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), K Nearest Neighbour (KNN), Xgboost (XGB), and Gradient Boosting (GB).

1) Bayesian Hyperparameter Optimization

Classification algorithms typically come with parameters that significantly impact their performance. These algorithms involve multiple hyperparameters that play a crucial role in determining the accuracy of the models. Therefore, it is vital to carefully fine-tune these hyperparameters, a process known as hyperparameter optimization. Traditionally, hyperparameter tuning has been viewed as more of an art than a precise science, relying on subjective judgments and iterative experimentation, which can be time-consuming [61]. While grid search represents a common hyperparameter optimization approach, it is computationally expensive and time-consuming, especially for problems with numerous hyperparameters [62] [63].

Bayesian optimization emerges as a systematic and efficient alternative, offering a more sophisticated solution to this challenge [64]. As evidenced by various studies [62] [65] [40], Bayesian optimization has demonstrated superior performance in various global optimization problems, particularly at hyperparameter tuning. It aligns with pursuing a more efficient and effective avenue for hyperparameter tuning in scenarios involving multiple hyperparameters. Therefore, the study leverages Bayesian hyperparameter optimization to fine-tune the hyperparameters of the classification models for stress classification. This approach allows the search for optimal hyperparameters to be systematic and can improve model performance.

TABLE 5. 2 Level stress: ANOVA test for EDA+IBI+BVP feature subset obtained using GA+MI feature selection algorithm

Features	dfb	dfw	F-Value	P-Value
Min HR	1	3764	2.77E+03	0.00E+00
mean_gsr	1	3764	1.53E+03	0.00e+00
Mean HR	1	3764	2.77E+03	0.00e+00
LF/HF	1	3764	7.45E+00	1.00E-14
RMSSD RR	1	3764	4.46E+01	3.05E-03
mean_scl	1	3764	2.77E+03	4.25E-04
slope_scl	1	3764	6.04E+01	6.36E-03
std_scr	1	3764	2.04E+01	6.51E-01
Mean BVP	1	3764	6.79E+00	9.69E-01

$*df_b$: degrees of freedom between classes, df_w: degrees of freedom within classes, p-value<0.05 indicate distinctive feature classes

TABLE 6. 3 Level stress: ANOVA test for EDA+IBI+BVP feature subset obtained using GA+MI feature selection algorithm

Features	dfb	dfw	F-Value	P-Value
Mean HR	2	5874	6.73E+04	0.00E+00
Min HR	2	5874	5.74E+03	0.00E+00
mean_gsr	2	5874	5.79E+03	0.00E+00
mean scl	2	5874	3.59E+03	3.10E-23
Power HF RR	2	5874	3.62E+03	4.42E-13
LF/HF	2	5874	3.59E+03	2.10E-04
RMSSD RR	2	5874	2.11E+03	1.25E-03
Mean BVP	2	5874	1.25E+03	3.90E-11
max_scr	2	5874	9.43E+02	3.08E-03
Slope scl	2	5874	1.25E+03	1.20E-05
STD RR/SDNN	2	5874	5.23E+01	2.87E-02

$*df_b$: degrees of freedom between classes, df_w: degrees of freedom within classes, p-value<0.05 indicate distinctive feature classes

III. EXPERIMENTATION AND RESULTS

A. PRELIMINARY ANALYSIS AND OPTIMIZATION

1. Statistical Feature Analysis: The study uses ANOVA to analyze the uniqueness of the optimal feature subset obtained during GA+MI feature selection. Tables 5 and 6 show the ANOVA results for binary and three-level stress classification, respectively.

Table 5 and Table 6 analysis reveal many features with statistically significant effects (p-value < 0.05) on differentiating between stress levels. Notably, features such as Min

TABLE 7. List of hyperparameters used for stress classification

Classifier	Hyperparameter	Value Range	Selected Value
RF	n_estimators	[10, 40, 60, 80, 100, 150, 200]	200
	max_depth	[2, 4, 8, 10, 12, 14]	4
	min_samples_split	[3, 5, 10, 15]	3
	random_state		0
DT	max_depth	[2, 4, 8, 10, 12, 14]	8
	min_samples_split	[2, 5, 10, 15, 20, 25]	10
KNN	n_neighbors	[3, 4, 5, 6, 7, 9, 12, 14, 16]	6
	metric	["euclidean", "manhattan", "minkowski"]	"minkowski"
	weights	["uniform", "distance"]	"distance"
	C	[0.1, 1, 10]	1
SVM	kernel	["linear", "rbf"]	"rbf"
	gamma	[0.01, 0.001, 0.0001, scale, auto]	0.1
XGB	n_estimators	[50, 100, 200, 500, 700]	500
	learning_rate	[0.01, 0.001, 0.1, 0.2]	0.1
	max_depth	[2, 4, 8, 10, 15]	8
GB	n_estimators	[100, 200, 500, 700, 900, 1000]	200
	learning_rate	[0.01, 0.001, 0.1, 0.2]	0.1
	max_depth	[2, 4, 8, 10, 12, 14]	4

HR, mean_gsr, and Mean HR consistently emerged as significant across both 2-level and 3-level classifications. Their high F-values and low p-values further support their strong influence on the model's performance. Additionally, Table 6 shows other features with significant effects, including LF/HF, Power HF RR, and STD RR/SDNN.

2. Performance Indexes: The performance of machine learning classifiers in stress classification systems can be assessed using a statistical approach called the "confusion matrix. This method helps determine how well the classifiers are doing. This work measures the performance using key metrics, including accuracy, F1 score, and AUC commonly used in evaluating machine learning models. This study uses positive and negative samples from the confusion matrix to evaluate the stress detection system's results. This matrix helps to calculate accuracy and F1 score using equation 8 and 9.

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \times 100 \tag{8}$$

$$F1Score = \frac{2TP}{2TP + FP + FN} \tag{9}$$

Additionally, the study calculates the model's global performance by measuring the Area Under the Curve (AUC). A value of 0.5 indicates a model with random performance, while 1 signifies flawless performance.

3. Hyperparameter Tuning: To optimize the performance of the classifiers, Bayesian hyperparameter optimization techniques are utilized to fine-tune hyperparameters. Table 7 lists the optimal hyperparameters utilized in the study. However, a thorough experimentation process is conducted to ensure the reliability and consistency of the computational results. This involves determining which hyperparameters are best left with their default values and which require tuning, employing the Bayesian technique.

B. CLASSIFICATION RESULTS

Prior research has yet to investigate the integration of ECG and EDA in the academic environment within the context of the MIST. Therefore, the current study explores the potential of wearable sensors, including IBI, BVP, and EDA biomarkers, in facilitating a comprehensive evaluation of stress responses.

The stress classification framework of the study primarily centers on identifying distinctive features from a complex, high-dimensional dataset through the application of the GA+MI Feature Selection Algorithm. Subsequently, several machine learning classifiers, namely RF, DT, SVM, KNN, XGBoost, and GB, are employed to classify these features for stress classification. Further, the study validates the computational results' reliability and stability through a ten-fold cross-validation method to calculate the average performance metrics.

Table 8, 9, 10 and 11 present the experimental results of 2-level and 3-level stress classification models, with and without the incorporation of the feature selection method. These results provide valuable insights into the efficacy of these models in accurately identifying different stress levels.

In 2-level stress classification, comparing individual physiological signals presents compelling results. Table 8 indicates that EDA and BVP features exhibit comparable performance without feature selection. Remarkably, IBI-derived HRV features exhibit the highest performance, achieving an accuracy of 86.54%.

Moreover, Table 9 presents the results after employing the GA+MI FSA. The analysis reveals a 3.49% increment in accuracy for EDA and a 1.77% increment for BVP. Further,

TABLE 8. Performance comparison of distinct physiological signals for 2-level stress classification with 58 original feature set

	RF			DT			SVM			KNN			XGB			GB		
Signals	**Acc**	F1	**AUC**	**Acc**	**F1**	**AUC**	**Acc**	**F1**	AUC	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**
EDA	76.4	77.69	84.4	76.32	75.61	78.3	55.7	55.69	60.2	60.15	59.57	69.67	78.93	77.91	87.3	**85.12**	85.64	94.63
IBI	83.84	83.62	89.42	80.37	79.11	83.65	78.72	78.88	86.52	80.5	80.33	83.98	82.92	81.4	86.83	**86.54**	86	93.02
BVP	83.91	83.92	88.54	78.61	78.75	78.37	54.7	55.03	64.29	78.83	78.73	86.66	82.88	83.97	88.29	**85.53**	85.91	91.4
EDA+IBI	93.35	92.54	97.03	90.22	90.47	95.03	83.96	83.5	87.1	83.5	81.45	87.4	92.94	92.98	96.7	**94.97**	95.35	97.86
IBI+BVP	89.76	89.67	95.02	86.22	87.14	92.64	79.71	80.78	84.98	83.88	83.75	89.75	90	90.1	95.69	**92.02**	92.56	96.94
EDA+BVP	88.3	88.13	93.76	87.37	87.35	91.54	62.15	62.03	67.7	80.91	82.55	87.32	88.21	88.16	94.72	**91.93**	91.91	97.38
EDA+IBI+BVP	93.37	92.43	97.83	90.54	90.87	94.15	84.54	83.21	89.26	84.59	85.63	90.41	92.23	91.67	97.36	**95.67**	95.73	98.59

TABLE 9. Performance comparison of distinct physiological signals for 2 level stress classification selected 9 features using GA-MI FSA

	RF			DT			SVM			KNN			XGB			GB		
Signals	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**
EDA	81.89	80.6	88.86	78.3	76.98	81.92	55.93	54.07	60.58	69.85	66.8	72.81	80.27	80.69	87.71	**88.61**	88.5	94.02
IBI	85.1	85.32	92.76	78.8	78.2	80.55	80.7	80.73	89.05	82.8	82.79	90.93	84.41	84.67	90.11	**90.86**	90.71	97.07
BVP	85.2	85.18	92.77	79	79.48	82.24	55.7	55.68	64.96	82.9	82.91	89.48	85.12	85.3	92.78	**87.3**	88.9	93.94
EDA+IBI	94.66	94.76	98.82	92.98	92.94	97.32	84.04	84.14	89.83	90.21	91.24	96.34	95.3	95.2	98.27	**97.07**	97.05	99.03
IBI+BVP	91.87	91.67	96.53	88.07	88.34	94.84	82.63	82.68	88.03	85.15	86.94	90.04	91.86	91.87	97.18	**94.83**	95.26	98.21
EDA+BVP	89.21	89.8	94.44	87.37	88.66	93.23	64.13	64.25	70.2	82.91	83.28	88.74	90.15	90.15	96.03	**93.42**	93.98	97.72
EDA+IBI+BVP	95.5	95.08	99.83	93.84	93.71	97.53	84.54	84	90.44	91.15	91.89	95.53	95.66	95.65	98.64	**98.28**	97.44	99.78

TABLE 10. Performance comparison of distinct physiological signals for 3-level stress classification with 58 original feature set

	RF			DT			SVM			KNN			XGB			GB		
Signals	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**
EDA	68.39	67.87	73.8	64.69	65.34	72.56	52.57	52.89	59.65	61.85	63.73	68.23	73.17	73.12	79.11	**75.01**	75.04	80.23
IBI	76.52	77.55	84.53	73.48	74.10	80.79	72.36	72.93	77.78	74.56	74.40	80.24	77.08	77.23	84.17	**79.74**	80.12	87.31
BVP	74.42	74.53	81.4	69.61	69.73	75.68	59.37	60.85	68.74	70.38	71.37	78.74	76.02	76.47	82.01	**79.70**	79.20	86.09
EDA+IBI	91.91	92.90	97.09	89.49	89.41	95.00	75.04	75.78	82.32	70.28	70.90	78.78	91.49	91.47	96.43	**93.87**	93.91	97.74
IBI+BVP	89.51	89.90	95.93	83.52	83.89	89.31	73.66	73.44	79.09	71.82	71.90	78.23	89.54	89.61	94.57	**91.92**	91.93	97.01
EDA+BVP	85.42	85.38	92.61	78.29	78.3	84.27	59.51	59.67	67.78	64.97	64.43	72.34	86.81	86.83	92.17	**89.43**	89.66	95.28
EDA+IBI+BVP	93.26	94.76	97.15	91.54	92.33	97.7	79.17	79.8	85.45	80.53	81.92	87.78	93.29	93.2	97.64	**95.16**	96.2	98.8

TABLE 11. Performance comparison of distinct physiological signals for 3-level stress classification with selected 11 features using GA-MI FSA

	RF			DT			SVM			KNN			XGB			GB		
Signals	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**	**Acc**	**F1**	**AUC**
EDA	71.6	70.32	77.98	67.07	68.34	73.3	52.52	52.78	59.78	67.13	67.42	72.26	75.17	75.27	81.56	**79.23**	79.98	87.76
IBI	79.93	79.27	89.48	75.88	75.56	84.84	75.26	75.08	83.1	77.3	77.63	86.04	80.04	80.41	87.72	**82.93**	82.91	89.48
BVP	77.15	77.13	83.17	70.83	70.85	77.66	64.31	64.18	72.67	72.23	73.01	80.34	78.63	78.85	85.3	**82.45**	83.46	88.11
EDA+IBI	93.12	94.23	98.21	91.76	91.98	97.78	76.68	76.23	83.32	73.07	73.87	79.77	93.16	93.28	97.16	**95.23**	96.05	98.47
IBI+BVP	90.43	90.35	96.95	84.98	84.65	92.02	75.11	75.74	81.54	73.65	73.25	79.67	90.06	90.4	95.82	**92.78**	92.24	97.63
EDA+BVP	88.48	88.35	93.7	80.62	80.57	86.45	62.98	62.62	68.96	77.77	77.92	84.21	89.03	89.02	95.6	**91.45**	91.42	96.42
EDA+IBI+BVP	95.65	96.56	99.52	93.95	94.83	98.4	80.61	80.67	87.67	89.16	90.57	95.42	95.66	95.75	98.13	**97.02**	96.53	99.13

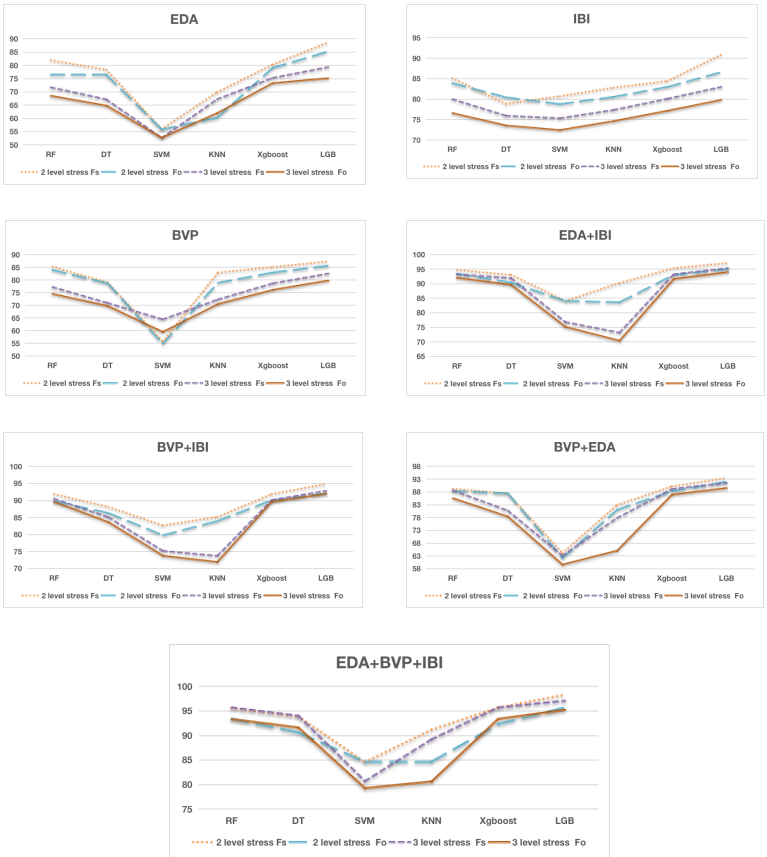

FIGURE 4. **Comparison of test accuracy of different classification models for 2-level and 3-level stress classification for the original feature set (F_o) and the selected feature set (F_s)**

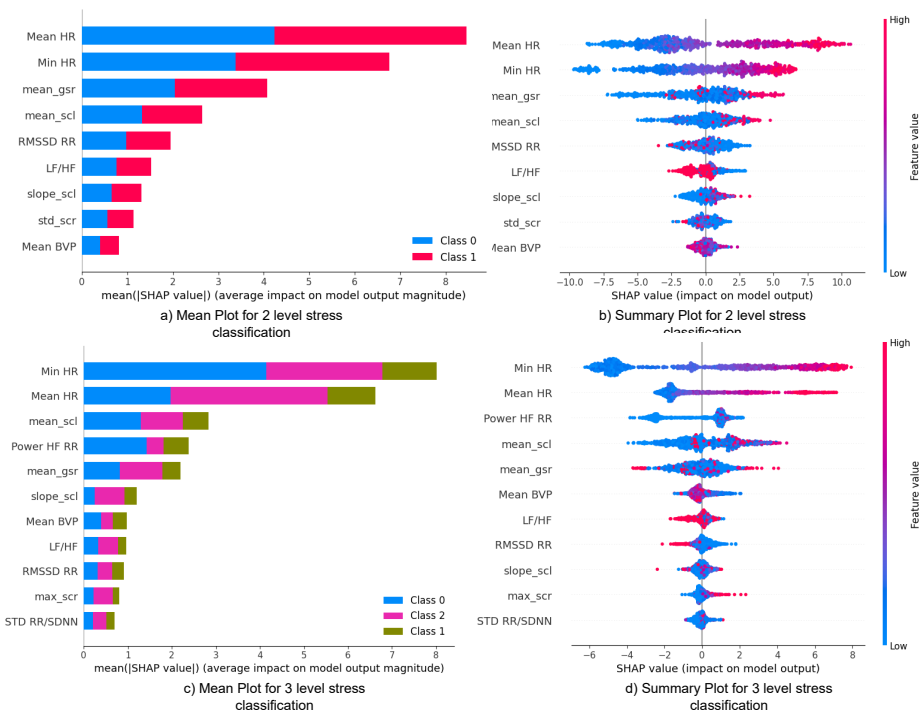

FIGURE 5. **Explaination of SHAP values**

the accuracy and F1 score of IBI-derived HRV features show significant improvements after feature selection, with an increment of 4.32% and 4.71%, respectively.

The combination of EDA and HRV features provides an accuracy of 94.97% without feature selection and reaches the highest accuracy of 97.07% with the feature selection algorithm. In contrast, incorporating all three features (EDA+HRV+BVP) results in a modest 2.68% accuracy increase, highlighting the substantial effectiveness of EDA and IBI signals in stress classification. Additionally, the study observes that RF, DT, and KNN demonstrate comparable performance, while XGB and GB consistently outperform other models with or without the feature selection algorithm GA+MI.

Similarly, Table 10 and 11 present the results before and after employing the feature selection method, showcasing enhanced results in the three-level stress classification task. The three-level classification models introduce an additional layer of complexity by distinguishing low, medium, and high stress levels. While individual physiological signals such as EDA and IBI display moderate performance, the combination of EDA and IBI outperforms combinations that involve BVP. However, using the GB classifier, the inclusion of BVP consistently achieves the highest accuracy, f1 score, and AUC score of 97.02%, 96.53%, and 97.47%, respectively.

The findings indicate that incorporating feature selection, particularly the GA+MI FSA, significantly enhances classification accuracy. This highlights the impact of eliminating redundant features from the original feature set, which has been found to have a negative impact on the model's performance. A visual representation of the feature selection comparison is shown in Figure 4, contrasting the outcomes between the original feature set, F_o, and the selected feature set, F_s, for both 2-level and 3-level stress classification.

C. EXPLAINABLE AI

Figure 5a and 5c depict bar plots presenting SHAP (Shapley Additive Explanations) values, offering a visual representation of the one-to-one scenario for feature importance in 2-level and 3-level stress classification, respectively. Feature importance is determined by calculating the mean absolute value for each feature across all provided samples. Figure 5b and Figure 5d illustrate summary plots of SHAP values for the top 9 features in the case of 2-level stress classification and the top 11 features for 3-level stress classification, utilizing the entire Stress dataset.

The summary plots illustrated in Figure 5b and Figure 5d show the impact and contribution of a specific feature to the stress classification model. The horizontal axis represents the SHAP values and their connection to high or low-stress predictions. The vertical axis above the zero point implies minimal impact on prediction results. For example, a SHAP value of zero suggests that the feature has no significant influence on predictions, while values close to zero indicate less influential predictions. In contrast, high-quality predictions, characterized by SHAP values significantly distant from zero, indicate positive or negative correlations. The figures also display how each feature impacts model output for varying stress levels. Each row in these plots corresponds to an individual feature, ordered by their mean absolute SHAP values. These values can be either negative or positive, reflecting the significance of each feature. These dots are stacked to reveal density and are colour-coded to indicate the intensity, with red signifying high feature value and blue indicating low feature value.

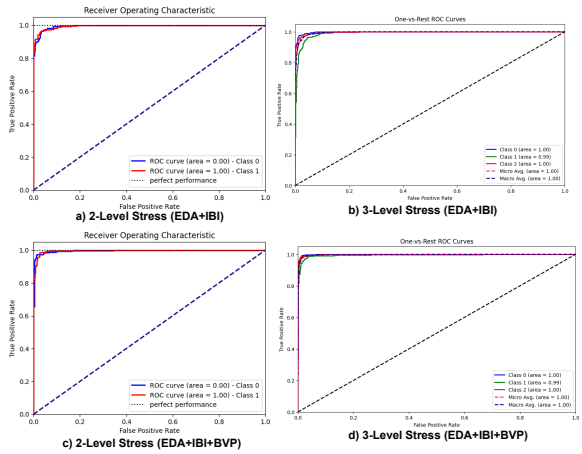

FIGURE 6. ROC of the best-evaluated results for integrated signals using Gradient Boosting

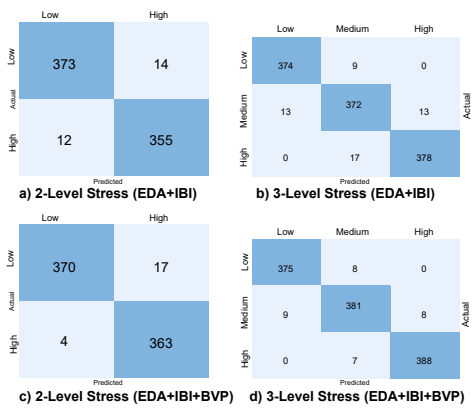

FIGURE 7. Confusion Matrix of the best-evaluated results for integrated signals using Gradient Boosting

Furthermore, Figure 5 shows that mean HR, Min HR, and mean_gsr are identified as highly important features for stress classification. This observation is consistent with the results of the ANOVA statistical analysis presented in Table 5 and 6. The low p-values and high F-score values in the ANOVA analysis support the conclusion regarding the significance of these features in predicting stress levels, providing a robust and reliable foundation for stress prediction.

IV. DISCUSSION

The experimental results underscore the efficacy of EDA and IBI as robust biomarkers for stress classification when employing the Gradient Boosting (GB) algorithm. In the context of two-level stress classification, the combination of EDA and IBI achieved a commendable accuracy of 97.07%. The subsequent addition of BVP features to the model (EDA+IBI+BVP) led to a marginal increase in accuracy, reaching 98.28%. This observation is reinforced by the Receiver Operating Characteristic (ROC) curves and confusion matrices presented in Figure 6 and 7, respectively. The figures depict comparable true positive (TP) and true negative (TN) rates and high Area Under the Curve (AUC) values, affirming the model's discriminatory power. Expanding the stress categorization to three levels incorporating BVP features further refined the model's performance, achieving accuracies and F1 scores of 97.02% and 96.53%, respectively. The confusion matrices in Figure 7 demonstrate the model's proficiency in distinguishing between the three stress levels, corroborated by the ROC curves in Figure 6, which exhibit a high AUC for each stress category. Notably, the EDA+IBI combination exhibited a comparable accuracy of 95.23% and an F1 score of 96.06%.

These results highlight the effectiveness of EDA and IBI in distinguishing between stress and non-stress levels. The marginal difference in accuracy observed upon including BVP features implies that IBI is derived from BVP and may contribute redundant information, resulting in only marginal improvement. It reinforces the notion that combining EDA and IBI can comprehensively capture a diverse range of physiological responses to stress, thereby ensuring a robust and reliable performance in stress classification.

Further, the study found that gradient boosting achieves better classification than XGBoost. XGBoost, known for its effectiveness in handling high-dimensional data with numerous features, might not be fully optimized when applied to datasets with limited features, as observed in this study. With a relatively smaller feature set, XGBoost might need a greater breadth of information to perform optimally. Therefore, gradient boosting, which is less sensitive to the number of features, emerged as a more suitable choice for this study.

This paper consists of stress detection and stress alleviation stages. The overall system shows an accuracy of 95.8% for stress classification. During the stress recovery phase, individuals listened to meditation audio, and the results indicate that this stage effectively aids in stress reduction.

The experimental findings underscore the significance of IBI (Interbeat Interval) and EDA (Electrodermal Activity) as primary contributors to stress classification. Figure 8 illustrates the Fast Fourier Transform (FFT) spectrum, obtained through Kubios analysis, displaying various stress levels. The FFT scrutinizes the HRV (Heart Rate Variability) signal derived from IBI. HRV serves as a measure of the time intervals between successive heartbeats. These intervals are influenced by the Autonomic Nervous System (ANS), which governs numerous automatic bodily functions, including heart rate, blood pressure, and respiration.

The ANS is divided into two distinct systems: the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). SNS is responsible for the "fight-or-flight" response, while PNS is responsible for the "rest-and-digest" response.

Figure 8 a) and 8 b) highlight a prominent peak in the low-frequency (LF) range associated with the activity of the sympathetic nervous system during moderate and high-stress situations. It also reveals a smaller peak in the high-frequency (HF) range, linked to parasympathetic nervous system activity. The figure indicates that even in the presence of stress, some parasympathetic activity continues, which may aid in alleviating the negative impacts of stress. Conversely, Figure 8 c) highlights a strong peak in the high-frequency (HF) range during low-stress situations, indicating increased activity in the parasympathetic nervous system.

Figure 9 shows the EDA graph to analyze EDA's phasic and tonic components. Several factors, including stress, arousal, and sweat, influence skin conductance. Phasic EDA refers to the quick changes in skin conductance that occur in response to specific stimuli or events. These changes are typically short-lived and are often associated with emotional arousal. Tonic EDA refers to the slower, more sustained changes in skin conductance that occur over time. Tonic EDA is thought to reflect the overall level of arousal and physiological activation.

The EDA graph is used to identify different stress states by analyzing EDA's phasic and tonic components. High stress shows a high concentration of phasic EDA responses and a high tonic EDA level. Moderate stress shows moderate phasic EDA responses and a moderate tonic EDA level. The figure shows the fall in phasic and tonic EDA levels, gradually decreasing EDA peaks for low-level stress.

A. COMPARISON WITH EXISTING WORK

The proposed work comprises stress detection and alleviation phases. The overall system achieves 98.28% and 97.02% accuracy for binary and 3-level stress classification. In the stress alleviation stage, participants listen to meditation audio. The study employs Explainable AI to identify influential features in stress detection.

Additionally, it investigates the impact of EDA and ECG on stress detection, analyzing the effects of stress and stress alleviation on the sympathetic and parasympathetic branches of ANS. The findings highlight the effectiveness of meditation audio in reducing stress within an academic setting.

Table 12 compares the previously proposed MIST studies with stress alleviation methods. It summarizes physiological signals, number of channels, subjects, stress alleviation technique, experiment duration, selected features, and results. Very few studies offer a solution for stress alleviation with MIST. Zhang et al. [20] classified stress levels with frequency domain features and employed the Fisher Ratio (FR) for feature reduction. The authors investigated distinct stress relief therapies such as watching and listening to alpha music,

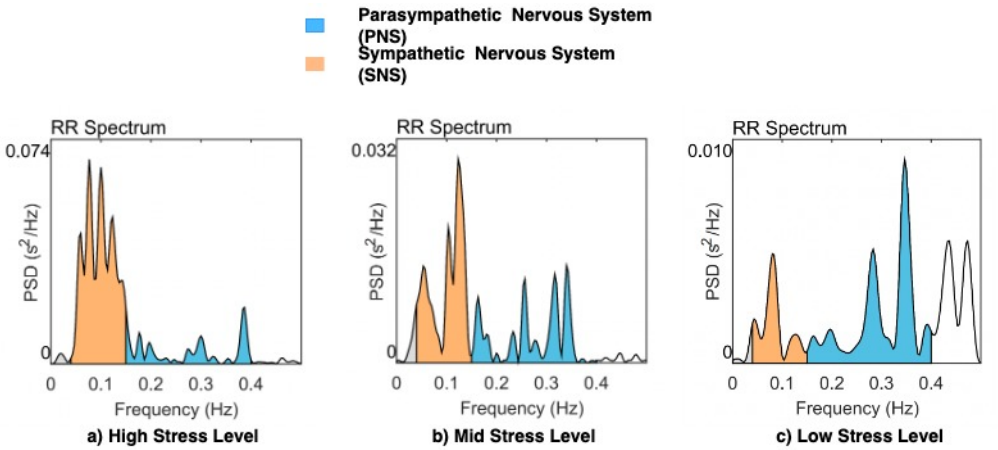

FIGURE 8. **Fast Fourier Transform Spectrum of distinct stress levels derived from IBI data and visualized using Kubios software: a) High-stress level b) Mid-stress level c) Low-stress level**

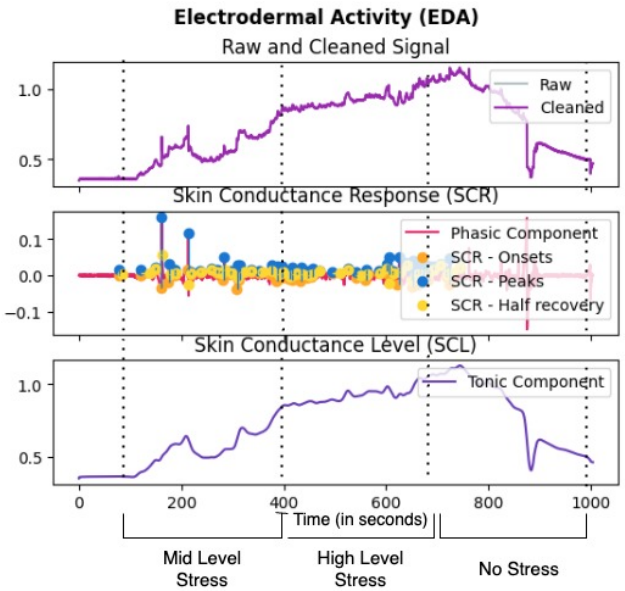

FIGURE 9. **EDA graph to analyze phasic and tonic components of EDA during stress classification**

videos of walking through a forest, bubble wrap popping, hugging a pillow, and squeezing a stress ball. The study found that hugging a pillow is the most effective method to mitigate stress, and the proposed stress alleviation technique works by comparing corresponding stress levels. Moreover, the study reported accuracy of 85.6 and 71.5 for 2-level stress and 3-level stress classification. Xia et al. [23] utilized EEG and ECG for stress detection and Fisher's discriminant ratio (FDR), minimum redundancy maximum relevance (mRmR), and information gain feature selection technique, achieving the highest accuracy of 79.5%. When the stress is induced via MIST, Sitting in a relaxed position and focusing on a circle on a computer screen is proposed for therapy. Moreover, the final selected features are not mentioned. In another study, Han et al. [25] utilized ECG and RESP and found sitting in a relaxed position to be stress-relieving therapy. However, sitting in an ideal relaxed position therapy can be influenced by environmental factors. External distractions or discomfort in the selected environment for therapy may hinder its efficacy. Sitting in a relaxed position is effective for the short term as some may find it difficult to induce stress or recall stressful situations genuinely.

Further, Setz et al. [24] proposed a general questionnaire and reading magazines as a stress relief method. However, reading a magazine for stress relief may vary in effectiveness due to individual preferences and the passive nature of the activity, which may only engage some equally. A quantitative evaluation of the effectiveness of the stress alleviation method needs to be presented in these studies.

The proposed work overcomes the limitations of the above studies. It employs an integrated feature selection algorithm, selecting only nine features and eleven features for 2-level- and 3-level stress classification. Further, it utilizes explainable AI to provide transparency in analysis. The detailed results show the significance of HRV and EDA features for stress detection systems. It also shows high classification accuracy using Bayesian optimization. Additionally, this study provides evidence that listening to meditation audio reduces stress levels, highlighting the effectiveness of such interventions in stress management.

V. CONCLUSION

Mental stress in academic environments adversely impacts students' well-being and academic performance. Through the exploration of wearable devices, particularly the utilization of IBI-derived HRV and electrodermal activity (EDA), the research demonstrates the potential for real-time stress detection and monitoring. The findings underscore the effectiveness of wearable technology in identifying stress among students, offering opportunities for timely interventions to alleviate its negative effects. Moreover, this paper has delved into the efficacy of stress reduction techniques, specifically the impact of listening to meditation audio. The results in-

TABLE 12. Performance comparison of proposed MIST studies with stress alleviation approach in the literature

Author	Signals	Channels	Sub	Stress Alleviation Technique	Experiment Duration	Feature Selection	Features	Classifiers	Accuracy	F1 Score	Other Metrics
Zhang et al. [20]	EEG	4	25	Watching a video, listening to alpha music, squeezing a stress ball, bubble wrap popping, and hugging a pillow.	16 min approx	Fisher Ratio	4	LDA	2 level: 85.6%, 3 level: 71.5%	–	–
Xia et al. [23]	EEG, ECG	128	22	Sitting in a relaxed position and focusing on a circle appearing on a computer screen	80 min	FDR, mRMR, Information Gain	–	SVM	79.45%	–	Recall: 81%, Specificity: 78%
Setz et al. [24]	EDA	–	32	Questionnaire and reading magazines	50 min	Not mentioned	16	LDA, SVM, NCC	82.80%	–	–
Han et al. [25]	ECG, RSP	–	39	Sitting in a relaxed position	18 minutes (excluding stress report filling time)	OOB	36	SVM, KNN, LDA, Adaboost	84%	83%	Recall: 84, Precision: 83
Proposed Work	ECG, EDA	–	30	Listening to meditation audio	22 min	GA + MI	9	RF, DT, SVM, KNN, GB, XGB	2 levels: 98.28%, 3 levels: 97.02%	2 levels: 97.44%, 3 levels: 96.53%	AUC: 99.78, AUC: 99.13

dicate that such practices can be valuable tools for stress reduction, providing students with accessible means to manage their stress levels. By incorporating a hybrid feature selection approach with machine learning classifiers, fine-tuned through Bayesian optimization has illustrated the effectiveness of the proposed classification models. This hybrid feature selection method improved the classification accuracy by 2.61% for two-level stress classification and by 1.86% for three-level stress classification compared to existing models without hybrid feature selection. This suggests incorporating HRV and EDA signals in wearable devices can enhance stress prediction accuracy. In the future, researchers can focus on further refining the utilization of HRV and EDA signals, which can contribute to developing more advanced and accurate wearable stress monitoring systems, thereby improving the overall effectiveness of stress management interventions.

CONFLICT OF INTEREST

The authors confirm that there are no known financial conflicts of interest or personal relationships that could have potentially influenced the findings presented in this paper.

DATA AVAILABILITY

Data will be made available on reasonable request.

REFERENCES

[1] P. S. Prabu, "A study on academic stress among higher secondary students," *International journal of humanities and social science invention*, vol. 4, no. 10, pp. 63–68, 2015.

[2] S. D. Ghatol, *Academic Stress among School Students*. Allied Publishers, 2019.

[3] A. Waqas, S. Khan, W. Sharif, U. Khalid, and A. Ali, "Association of academic stress with sleeping difficulties in medical students of a Pakistani medical school: a cross sectional survey," *PeerJ*, vol. 3, p. e840, 2015.

[4] V. M. Bhujade, "Depression, anxiety and academic stress among college students: A brief review.," *Indian Journal of Health & Wellbeing*, vol. 8, no. 7, 2017.

[5] S. Maji, A. Chaturmohta, D. Deevela, S. Sinha, S. Tarsolia, and A. Barsaiya, "Mental health consequences of academic stress, amotivation, and coaching experience: A study of india's top engineering undergraduates," *Psychology in the Schools*.

[6] S. Byun, A. Y. Kim, E. H. Jang, S. Kim, K. W. Choi, H. Y. Yu, and H. J. Jeon, "Detection of major depressive disorder from linear and nonlinear heart rate variability features during mental task protocol," *Computers in biology and medicine*, vol. 112, p. 103381, 2019.

[7] M. Jafari, A. Shoeibi, M. Khodatars, S. Bagherzadeh, A. Shalbaf, D. L. García, J. M. Gorriz, and U. R. Acharya, "Emotion recognition in EEG signals using deep learning methods: A review," *Computers in Biology and Medicine*, p. 107450, 2023.

[8] L. Holtz, M. Martinez, K. Paton, K. Rosich, and E. Schnittka, "Effects of physiological stress response on short-term memory recall," 2017.

[9] C. Anders and B. Arnrich, "Wearable electroencephalography and multimodal mental state classification: A systematic literature review," *Computers in Biology and Medicine*, p. 106088, 2022.

[10] A. Arsalan and M. Majid, "Human stress classification during public speaking using physiological signals," *Computers in biology and medicine*, vol. 133, p. 104377, 2021.

[11] M. Ullah, S. Akbar, A. Raza, and Q. Zou, "DeepAVP-TPPred: identification of antiviral peptides using transformed image-based localized descriptors and binary tree growth algorithm," *Bioinformatics*, vol. 40, no. 5, p. btae305, 2024.

[12] S. Akbar, Q. Zou, A. Raza, and F. K. Alarfaj, "iAFPs-Mv-BiTCN: Predicting antifungal peptides using self-attention transformer embedding and transform evolutionary based multi-view features with bidirectional temporal convolutional networks," *Artificial Intelligence in Medicine*, vol. 151, p. 102860, 2024.

[13] S. Akbar, A. Raza, and Q. Zou, "Deepstacked-AVPs: predicting antiviral peptides using tri-segment evolutionary profile and word embedding based multi-perspective features with deep stacking model," *BMC bioinformatics*, vol. 25, no. 1, p. 102, 2024.

[14] A. Raza, J. Uddin, A. Almuhaimeed, S. Akbar, Q. Zou, and A. Ahmad, "AIPs-SnTCN: Predicting anti-inflammatory peptides using fasttext and transformer encoder-based hybrid word embedding with self-normalized temporal convolutional networks," *Journal of chemical information and modeling*, vol. 63, no. 21, pp. 6537–6554, 2023.

[15] S. Akbar, A. Raza, T. Al Shloul, A. Ahmad, A. Saeed, Y. Y. Ghadi, O. Mamyrbayev, and E. T. Eldin, "pAtbP-EnC: identifying anti-tubercular peptides using multi-feature representation and genetic algorithm based deep ensemble model," *IEEE Access*, 2023.

[16] A. Kumar, K. Sharma, and A. Sharma, "Hierarchical deep neural network for mental stress state detection using IoT based biomarkers," *Pattern Recognition Letters*, vol. 145, pp. 81–87, 2021.

[17] Y. S. Can, N. Chalabianloo, D. Ekiz, and C. Ersoy, "Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study," *Sensors*, vol. 19, no. 8, p. 1849, 2019.

[18] B. Egilmez, E. Poyraz, W. Zhou, G. Memik, P. Dinda, and N. Alshurafa, "UStress: Understanding college student subjective stress using wrist-based passive sensing," in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pp. 673–678, IEEE, 2017.

[19] K. Dedovic, R. Renwick, N. K. Mahani, V. Engert, S. J. Lupien, and J. C. Pruessner, "The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain," *Journal of Psychiatry and Neuroscience*, vol. 30, no. 5, pp. 319–325, 2005.

[20] Y. Zhang, Q. Wang, Z. Y. Chin, and K. K. Ang, "Investigating different stress-relief methods using electroencephalogram (EEG)," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 2999–3002, IEEE, 2020.

[21] F. Al-Shargie, "Prefrontal cortex functional connectivity based on simultaneous record of electrical and hemodynamic responses associated with mental stress," *arXiv preprint arXiv:2103.04636*, 2021.

[22] J. Minguillon, E. Perez, M. A. Lopez-Gordo, F. Pelayo, and M. J. Sanchez-Carrion, "Portable system for real-time detection of stress level," *Sensors*, vol. 18, no. 8, p. 2504, 2018.

[23] L. Xia, A. S. Malik, and A. R. Subhani, "A physiological signal-based method for early mental-stress detection," *Biomedical Signal Processing and Control*, vol. 46, pp. 18–32, 2018.

[24] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Transactions on information technology in biomedicine*, vol. 14, no. 2, pp. 410–417, 2009.

[25] L. Han, Q. Zhang, X. Chen, Q. Zhan, T. Yang, and Z. Zhao, "Detecting work-related stress with a wearable device," *Computers in Industry*, vol. 90, pp. 42–49, 2017.

[26] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *Journal of biomedical informatics*, vol. 73, pp. 159–170, 2017.

[27] L. Zhu, P. Spachos, P. C. Ng, Y. Yu, Y. Wang, K. Plataniotis, and D. Hatzinakos, "Stress Detection Through Wrist-Based Electrodermal Activity Monitoring and Machine Learning," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[28] C.-P. Hsieh, Y.-T. Chen, W.-K. Beh, and A.-Y. A. Wu, "Feature selection framework for xgboost based on electrodermal activity in stress detection," in *2019 IEEE International Workshop on Signal Processing Systems (SiPS)*, pp. 330–335, IEEE, 2019.

[29] P. Siirtola, "Continuous stress detection using the sensors of commercial smartwatch," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pp. 1198–1201, 2019.

[30] D. Jaiswal, D. Chatterjee, R. Gavas, R. K. Ramakrishnan, and A. Pal, "Person and Stressor Independent Generic Model for Stress Detection Using GSR," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 7195–7198, IEEE, 2021.

[31] T. Androutsou, S. Angelopoulos, E. Hristoforou, G. K. Matsopoulos, and D. D. Koutsouris, "Automated Multimodal Stress Detection in Computer Office Workspace," *Electronics*, vol. 12, no. 11, p. 2528, 2023.

[32] S. Campanella, A. Altaleb, A. Belli, P. Pierleoni, and L. Palma, "A Method for Stress Detection Using Empatica E4 Bracelet and Machine-Learning Techniques," *Sensors*, vol. 23, no. 7, p. 3565, 2023.

[33] A. Hemakom, D. Atiwiwat, and P. Israsena, "ECG and EEG based machine learning models for the classification of mental workload and stress levels for women in different menstrual phases, men, and mixed sexes," *Biomedical Signal Processing and Control*, vol. 95, p. 106379, 2024.

[34] M. Benchekroun, P. E. Velmovitsky, D. Istrate, V. Zalc, P. P. Morita, and D. Lenne, "Cross dataset analysis for generalizability of HRV-based stress detection models," *Sensors*, vol. 23, no. 4, p. 1807, 2023.

[35] M. Quiñones, A. Van den Broeck, and H. De Witte, "Do job resources affect work engagement via psychological empowerment? A mediation analysis," *Revista de Psicología del Trabajo y de las Organizaciones*, vol. 29, no. 3, pp. 127–134, 2013.

[36] H. F. Posada-Quintero, J. P. Florian, A. D. Orjuela-Cañón, T. Aljama-Corrales, S. Charleston-Villalobos, and K. H. Chon, "Power spectral density analysis of electrodermal activity for sympathetic function assessment," *Annals of biomedical engineering*, vol. 44, pp. 3124–3135, 2016.

[37] A. Greco, G. Valenza, J. Lazaro, J. M. Garzon-Rey, J. Aguilo, C. de la Cámara, R. Bailón, and E. P. Scilingo, "Acute stress state classification based on electrodermal activity modeling," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 788–799, 2021.

[38] Y. Liu and S. Du, "Psychological stress level detection based on electrodermal activity," *Behavioural brain research*, vol. 341, pp. 50–53, 2018.

[39] C. Schubert, M. Lambertz, R. Nelesen, W. Bardwell, J.-B. Choi, and J. Dimsdale, "Effects of stress on heart rate complexity—a comparison between short-term and chronic stress," *Biological psychology*, vol. 80, no. 3, pp. 325–332, 2009.

[40] R. Turner, D. Eriksson, M. McCourt, J. Kiili, E. Laaksonen, Z. Xu, and I. Guyon, "Bayesian optimization is superior to random search for machine learning hyperparameter tuning: Analysis of the black-box optimization challenge 2020," in *NeurIPS 2020 Competition and Demonstration Track*, pp. 3–26, PMLR, 2021.

[41] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, "Hyperparameter optimization for machine learning models based on bayesian optimization," *Journal of Electronic Science and Technology*, vol. 17, no. 1, pp. 26–40, 2019.

[42] S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," *Journal of health and social behavior*, pp. 385–396, 1983.

[43] R. Champseix, L. Ribiere, and C. Le Couedic, "A Python Package for Heart Rate Variability Analysis and Signal Preprocessing.," *Journal of Open Research Software*, vol. 9, no. 1, 2021.

[44] S. Föll, M. Maritsch, F. Spinola, V. Mishra, F. Barata, T. Kowatsch, E. Fleisch, and F. Wortmann, "FLIRT: A feature generation toolkit for wearable data," *Computer Methods and Programs in Biomedicine*, vol. 212, p. 106461, 2021.

[45] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures.," *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[46] G. Cosoli, A. Poli, L. Scalise, and S. Spinsante, "Measurement of multimodal physiological signals for stimulation detection by wearable devices," *Measurement*, vol. 184, p. 109966, 2021.

[47] B. Remeseiro and V. Bolon-Canedo, "A review of feature selection methods in medical applications," *Computers in biology and medicine*, vol. 112, p. 103375, 2019.

[48] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction," *Journal of Applied Science and Technology Trends*, vol. 1, no. 2, pp. 56–70, 2020.

[49] M. M. Kabir, M. Shahjahan, and K. Murase, "A new local search based hybrid genetic algorithm for feature selection," *Neurocomputing*, vol. 74, no. 17, pp. 2914–2928, 2011.

[50] I.-S. Oh, J.-S. Lee, and B.-R. Moon, "Hybrid genetic algorithms for feature selection," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 26, no. 11, pp. 1424–1437, 2004.

[51] L. I. Kuncheva and L. C. Jain, "Nearest neighbor classifier: Simultaneous editing and feature selection," *Pattern recognition letters*, vol. 20, no. 11-13, pp. 1149–1156, 1999.

[52] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 2, pp. 44–49, 1998.

[53] M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn, and A. K. Jain, "Dimensionality reduction using genetic algorithms," *IEEE transactions on evolutionary computation*, vol. 4, no. 2, pp. 164–171, 2000.

[54] M. Mitchell, "An introduction to genetic algorithms mit press," *Cambridge, Massachusetts. London, England*, vol. 1996, 1996.

[55] J. J. Grefenstette, "Genetic algorithms and machine learning," in *Proceedings of the sixth annual conference on Computational learning theory*, pp. 3–4, 1993.

[56] D. Ramdania, M. Irfan, F. Alfarisi, and D. Nuraiman, "Comparison of genetic algorithms and Particle Swarm Optimization (PSO) algorithms in course scheduling," in *Journal of Physics: Conference Series*, vol. 1402, p. 022079, IOP Publishing, 2019.

[57] M. Rostami, K. Berahmand, and S. Forouzandeh, "A novel community detection based genetic algorithm for feature selection," *Journal of Big Data*, vol. 8, no. 1, pp. 1–27, 2021.

[58] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern recognition*, vol. 33, no. 1, pp. 25–41, 2000.

[59] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Transactions on neural networks*, vol. 5, no. 4, pp. 537–550, 1994.

[60] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural computing and applications*, vol. 24, pp. 175–186, 2014.

[61] P. Koch, O. Golovidov, S. Gardner, B. Wujek, J. Griffin, and Y. Xu, "Autotune: A derivative-free optimization framework for hyperparameter tuning," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 443–452, 2018.

[62] M. R. Hossain and D. Timmer, "Machine learning model optimization with hyper parameter tuning approach," *Glob. J. Comput. Sci. Technol. D Neural Artif. Intell*, vol. 21, no. 2, 2021.

[63] S. Andradóttir, "A review of random search methods," *Handbook of Simulation Optimization*, pp. 277–292, 2014.

[64] V. Nguyen, "Bayesian optimization for accelerating hyper-parameter tuning," in *2019 IEEE second international conference on artificial intelligence and knowledge engineering (AIKE)*, pp. 302–305, IEEE, 2019.

[65] F. Fatih, Z. En-Naimani, and K. Haddouch, "Comparative Study of Bayesian Optimization Process for the Best Machine Learning Hyperparameters," in *International Conference On Big Data and Internet of Things*, pp. 239–249, Springer, 2022.

SHIKHA received B.Tech degree in Information Technology from the Guru Gobind Singh Indraprastha University (GGSIPU) Delhi in 2014 and an M.Tech degree in Computer Science and Engineering from Indira Gandhi Delhi Technical University (IGDTUW) Delhi in 2016. She is pursuing a doctorate of philosophy from Delhi Technological University(DTU), Delhi. Her research interests are Human-computer Interaction in healthcare and Machine Learning.

DR. DIVYASHIKHA SETHIA received her B.Tech degree in Computer Science and Engineering from the Maharaja Sayajirao University of Baroda, India, in 1997, and her M.Tech degree in the same field from the Indian Institute of Technology Delhi, India, in 2006, with a focus on distributed systems and mobile computing. She has over five years of industry experience in the software development sector. Since 2010, she has been an Assistant Professor in the Department of Computer Science and Engineering at Delhi Technological University (DTU) Delhi. Her research interests encompass telemedicine, security in distributed systems, and the application of brain-computer interfaces in healthcare. She is the author of more than 13 journal publications and 50 conference papers.

PROF. S. INDU (Senior Member, IEEE) received the B.Tech. degree from the TKM College of Engineering, University of Kollam, India, in 1987, the M.Tech degree from the College of Engineering Trivandrum, University of Kerala, Thiruvananthapuram, India, in 1990, and a Ph.D. degree from the University of Delhi, India, in 2012. She joined the Electronics and Communication Engineering Department at Delhi College of Engineering, Delhi, in 1999. She works as Dean (student Welfare) at DTU and a Professor of the ECE Department at Delhi Technological University (DTU), Delhi. Her research interests include computer vision, sensor networks, and image processing. Prof. Indu received the Commendable Research Award from DTU in 2018 and 2019. She has over 150 publications in books, book chapters, and peer-reviewed journals.