Article

Enhancing EEG-Based Mental Stress State Recognition using an Improved Hybrid Feature Selection Algorithm

Ala A. Hag [1, *]**, Dini Handayani** [1,*]**, Maryam Altalhi** [2,*]**, Thulasyammal Pillai** [1]**, Teddy Mantoro** [3]**, Mun Hou Kit** [4]**, and Fares Al-Shargie** [5,*]

[1]   School of Computer Science & Engineering, Taylor's University, Jalan Taylors, 47500, Subang Jaya, Selangor, Malaysia; dinioktarina.dwihandayani@taylors.edu.my

[2]   Department of Management Information System, College of Business Administration, Taif University, P.O. BOX 11099, Taif 21944, Saudi Arabia; marem.m@tu.edu.sa

[3]   Faculty of Engineering and Technology, Sampoerna University, Jakarta 12780, Indonesia; teddy.mantoro@sampoernauniversity.ac.id

[4]   Department of Mechatronic and Biomedical Engineering, Universiti Tunku Abdul Rahman, 43000 Bandar Sungai Long, Malaysia; munhk@utar.edu.my

[5]   Department of Electrical Engineering, American University of Sharjah, UAE

*   Correspondence: Maryam Altalhi (email:marem.m@tu.edu.sa), Dini Handayani (email: dinioktarina.dwihandayani@taylors.edu.my), Fares Al-Shargie (email:fyahya@aus.edu); Ala Hag (email:alaahmedyahyahag@sd.taylors.edu.my);

Abstract: Mental stress state recognition using electroencephalogram (EEG) signals for real-life applications needs a conventional wearable device. This requires an efficient number of EEG channels and an optimal feature set. The main objective of the study is to identify an optimal feature subset that can best discriminate mental stress states while enhancing the overall performance. Thus, multi-domain feature extraction methods were employed, namely, time domain, frequency domain, time-frequency domain, and network connectivity features, to form a large feature vector space. To avoid the computational complexity of high dimensional space, a hybrid feature selection (FS) method of minimum Redundancy Maximum Relevance with Particle Swarm Optimization and Support Vector Machine (mRMR-PSO-SVM) is proposed to remove noise, redundant, and irrelevant features and keep the optimal feature subset. The performance of the proposed method is evaluated and verified using four datasets, namely EDMSS, DEAP, SEED, and EDPMSC. To further consolidate, the effectiveness of the proposed method is compared with that of the state-of-the-art heuristic methods. The proposed model has significantly reduced the features vector space by an average of 70% in comparison to the state-of-the-art methods while significantly increasing overall detection performance.

Keywords: brain-computer interface (BCI); electroencephalography (EEG); stress state recognition; feature selection; particle swarm optimization (PSO); mRMR; SVM; DEEP; SEED

1. Introduction

Feature selection (FS) is a crucial processing step in machine learning that leads to the development of efficient real-world applications. FS aims to identify the most relevant feature sets for a given task by removing irrelevant and redundant features. Hence, it reduces the high dimensionality space and prevents the incidence of the over-fitting problem caused by the curse of dimensionality [1,2]. In particular, FS approaches have demonstrated their significant impact on improving the overall classification performance of a given problem, in terms of the quality of the extracted features and in reducing the computational costs [3]. Therefore, current state-of-art efforts employ FS as an optimization technique to enhance the classification performance by selecting the optimal feature set [2]. Furthermore, FS has been successfully utilized to solve several classification problems from various domains, for instance, data mining [2], pattern recognition [4–6], and other domains where the high dimensionality occurred.

Specifically, pattern recognition such as Electroencephalography (EEG) is a vulnerable domain requiring the extraction of relevant patterns from high dimensional space. Current research studies employed EEG to acquire brain activities because it is reliable, affordable, portable, and provides high temporal resolution of the brain signals [7,8]. In multi-channel EEG, several features from the time domain, frequency domain, time-frequency domain, spatial domain, etc., contribute to the high dimensional feature space in which one aims to recognize or assess several brain states such as seizure detection (epilepsy) [9], motor imaginary [10], depression [11], emotion detection [12,13], and mental stress recognition [14]. Recently, EEG signals have been used extensively in the field of emotion recognition, particularly in the recognition of distress due to its harmful influence on physical and mental health [15,16]. However, one of the major challenges in building a successful model for stress detection is finding the most appropriate features. Due to that researchers employ several features, extracted from time domain, frequency domain, brain connectivity network or time-frequency network, and a combination of one or more methods, to find their association with mental stress [16].

Despite the great efforts from community researchers, in recent studies, for enhancing the classification accuracy of mental stress state recognition using EEG signals, few significant studies utilized multi-feature domains and multi-channel EEG with feature selection method. Yet, there is no solid conclusion of what is the optimal feature subset for stress recognition [4,5,17].

Consequently, current researchers acknowledge that multi-feature and multi-channel analyses are required to establish informative feature space in which a good interpreter can eventually produce effective alarms of the occurrence of mental state [5,12,17–19], allowing subjects to seek appropriate treatment at an early stage. The success in finding an optimal feature set for stress detection would be an essential step toward the creation of real-world applications that would provoke clinical or behavioral intervention if stress continued to worsen [20]. However, extracting EEG multi-features from different domains would result in high dimensionality that may contain irrelevant features that are not helpful for machine learning classification due to the enormous search space known as the "curse of dimensionality" [1]. Thus, FS becomes an essential pipeline to address these problems in EEG-specific domain' analysis by selecting an optimal feature subset and reducing system complexity [1,21].

The feature selection approaches are often categorized into two main types: filters, and wrappers. Some other selection approaches are discussed in various literature [1,22]. Filter methods measure the degree of the importance between each feature without the involvement of subsequent learning algorithms [23]. In contrast, wrapper methods rely on prediction models (e.g., SVM, KNN) to estimate the importance of features via classification algorithm evaluation [2]. In comparison, filters methods usually outperform wrappers in terms of computational speed because they use statistical measurements between features such as feature distance, information gain, and feature dependency; however, wrappers methods proved to be better at finding the importance of feature subsets that improve overall classification performance [23,24]. Nonetheless, these feature selection methods still suffer from some drawbacks, such as the feature selected subset can present a correlation between features (redundancy), be trapped in local optima, and may trigger a high computational cost [25,26]. Also, they tend to perform global searches to find the optimal features, yet it is impossible in most cases [27].

To address the feature selection challenges, an efficient global search technique is needed to select a (near) optimal feature subset from the original feature set. With global searchability, the swarm intelligence-based heuristic search methods have shown superior performance to obtain optimal solutions [28]. Metaheuristics are also seen to be superior to random searches because, in the worst-case scenario, they can perform as a comprehensive search method. Swarm intelligence-based (SI) heuristic search methods aim to investigate the behavior of a group of agents in self-organized communities, such as ants, bees, moths, and birds [1]. Recently, several SI algorithms, such as the Ant Colony

Optimization (ACO) [28], Bee Optimization Algorithm (BeOA) [29], Moth-flame optimizer (MFO) [30], Multi-Verse Optimizer (MVO) [31], Butterfly Optimization algorithm (BOA) [32], Bat Algorithm (BAT)[33], Firefly Algorithm (FFA) [1], Grey Wolf Optimizer (GWO) [34], Moth Optimization Algorithm (MOA) [30], Whale Optimization Algorithm (WOA) [2] and Particle Swarm Optimization (PSO) [2,35] have been successfully utilized to discover the optimal feature subset. However, despite the excellent findings, most of these algorithms have a poor convergence rate and are entrapped in local optima [32]. Recently, researchers developed various hybrid algorithms using swarm intelligence models, which fused at least two approaches to improve each method's performance and overcome challenges occurred in search space [1,36].

PSO is a relatively recent optimization technique that is more computationally efficient than the existing metaheuristic methods. The method has shown to be a valuable solution for optimization issues due to its effectiveness and ease of implementation. However, there are some major issues with the conventional PSO when used for feature selection, such as lack of exploitation for particular problems [37]. Moreover, according to the "NO Free Lunch" (NFL) theorem, no feature selection algorithm can handle all optimization problems. As a result, if one algorithm outperforms other algorithms in one specific domain problem, it may not exceed them in another. Furthermore, the standard PSO is suggested for the continuous optimization problem, which cannot be employed for feature selection issues with binary solution space [37]. Therefore, the considerations mentioned above motivate us to propose a new hybrid method to make it more suited for feature selection and to effectively balance exploration and exploitation in the search process.

In this work, a hybrid of minimum-Redundancy Maximum Relevance and PSO (mRMR-PSO) feature selection methods are proposed. The proposed method aims to improve the exploitation of the PSO algorithm. The mRMR is utilized to enhance the exploitation of PSO, as a building block of PSO. It ranks a set of features by applying Pearson's correlation to minimize the redundancy among the subset of features while maximizing the features' relevance using the f-test. The output of the selected optimal features is then classified using a support vector machine (SVM).

In summary, the main contributions of this work are given as follows:
1. Develop an experimental protocol to induce stress on participants while solving mental arithmetic tasks under time pressure and negative feedback.
2. Extract multi-domain features from multi-EEG channels and fused them to form a large pool of feature vectors.
3. Propose a novel EEG feature selection method called mRMR-PSO-SVM to improve the search of local optimal and fit for binary feature selection.
4. Validated the proposed method by utilizing our dataset with another three public datasets of EEG on mental stress state and compared its performance with several metaheuristic algorithms.

The structure of the paper is organized as follows. Section 2, explains the datasets structures, preprocessing, and feature extraction as well as presenting the methodology of the proposed framework. In section 4, the results of the study are presented. Finally, the discussion and conclusion are provided in sections 5 and 3, respectively.

2. Experiment and Materials

In this study, our EEG Dataset for Mental Stress State (EDMSS) and three other public datasets were utilized to validate the proposed method. The below subsections describe the details for each dataset.

2.1. Participants

Twenty-two healthy right-handed males (aged 26± 4 with a head size of 56± 2 cm) participated in this experiment. The experiment was conducted between 3.00 and 4.30 p.m to avoid circadian rhythm influences on the alpha-amylase measurement. These

individuals had no history of psychiatric, neurological, or psychotropic drug use. The participants were placed in a comfortable room with good air conditioning and instructed to avoid moving their heads as possible throughout the experiment to avoid any environmental stress. All participants were informed about the study and signed consent forms before the trial began. The experiment procedures followed the Declaration of Helsinki and ethics approval granted by the local ethics review committee at Universiti Teknologi PETRONAS.

2.2. Stress Inducement Method

We induced stress by utilizing mental arithmetic tasks with time pressure and negative feedback as reported in our previous studies [20,38,39]. Then we carried out the experimental protocol in four steps.

Step 1: The participants were given a brief introduction to the tasks that would be assigned to them and collected an alpha-amylase sample (S1) as a baseline.

Step 2 (training phase): Each of the participants practiced the mental arithmetic (MA) task for five minutes to estimate how long it would take to answer each question. The MA task involved three single-digit integers (range from 0 to 9) and used two operators; plus (+) or minus (–), (for example, 7-3 + 1), see Figure 1. Each question's response was displayed on a computer monitor in the order of '0' to '9,' and the participant had to select the correct answer with a single left click on the mouse.

Step 3 (stress phase): Participants performed the same MA task under time pressure and negative emotional feedback. In this context, the allocated time to answer the MA task was reduced by 10% compared to the average time taken to answer each question during phase 2. Participants who answered incorrectly or did not submit their answer within the allocated time received negative comments in the form of a notification display in the monitor as "Incorrect" or "Time's up", "correct" alongside the average performance of participants and Excellent user.

Step 4 (rest phase): Participants were asked to look at a fixation cross on the computer with a black background to sustain their attention to the monitor. The entire experiment lasted about 15 minutes including the experimental setup. Meanwhile, the task presentation during the stress and rest states lasted for five minutes. In this experiment, we displayed the MA task in a block design following the fNIRS protocol [40,41].

The task's block architecture is shown in Figure 1. Each block began with a 30-seconds MA stress task, followed by a 20-rest task. The baseline was taken for a total of 20 seconds before starting the experiment. Immediately, at the end of the five blocks of the MA task, we collected the second sample of alpha-amylase (S2) as shown in Figure 1. To check if the participants were paying attention to the exercise, the accuracy with which they answered the questions was calculated. The average score was 95% accuracy in the training phase and 40% in the stress phase, as reported in the original published MIST paper [38].
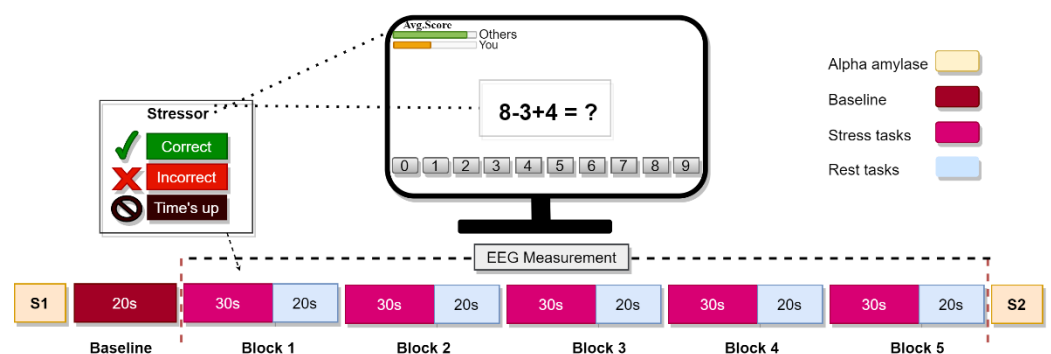

Figure 1 Experiment block design. A total of five blocks for the stress and rest tasks. In each block, arithmetic tasks were generated for 30 seconds followed by 20 seconds' rest. Alpha-amylase samples

(S1-S2) were taken five minutes before the experiment began as a baseline and five minutes after the experiment ended.

2.3. Data Acquisition

Brain activity was recorded using EEG signal from prefrontal cortex using seven electrodes namely: FP1, FP2, F3, F4, Fz, F7, and F8, plus two reference electrodes A1 and A2 placed to the earlobes as shown in Figure **2**. The BrainMaster 24E system is an EEG machine with wet electrodes utilized in this study with a sampling rate of 256Hz. However, for salivary alpha-amylase, we measure the cortisol activities using a hand-held monitor called COCORO meter (Nipro, Osaka, Japan).

Figure 2 EEG Electrodes placement on the scalp.

2.4. Description of Public Datasets

Three publicly available datasets were also used in this study for validation of the proposed method. The summary of the dataset's contents data used in this study is shown in Table 1. The description for each of the three datasets are given below:

2.4.1. DEAP Dataset

A Database for Emotion Analysis using Physiological Signals (DEAP) is a well-known publicly available dataset for emotion classification [42]. The DEAP dataset contains multiple physiological signals for the evaluation of emotions. Thirty-two healthy participants have participated in the experiment. The EEG signals were acquired with 32
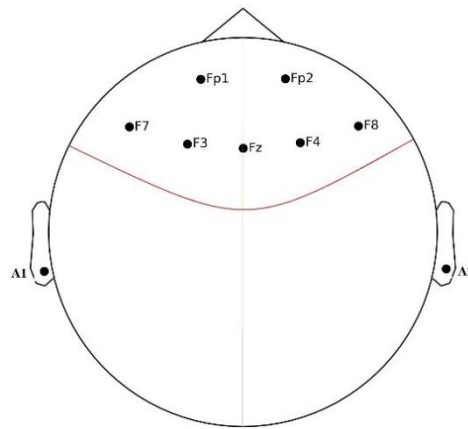
channels/electrodes while watching music videos. A total of 40 different music videos, each 60 seconds long for emotional stimulation. The signals were then downsampled to 128 Hz and preprocessed from artifacts and noise. The EEG signals were cleaned from EOG artifacts, de-noised using bandpass filters from 4-45hz, and then data was averaged to a common reference. In the DEAP dataset, the emotional state was labeled based on arousal and valence of self-assessment manikins (SAM) [43]. The EEG signals were annotated based on the online self-assessment rating SAM scale provided by DEAP for valence and arousal. Therefore, in this study, the valence and arousal values were considered to identify calm and stress tasks for each participant using Equations (1) and (2), derived from [17,44]:

$$Calm = (arousal < 4) \cap (4 < valence < 6) \qquad (1)$$

$$Stress = (arousal > 5) \cap (valence < 3) \qquad (2)$$

By applying the rules of selecting stress and calm states from each participant, a result of 25 participants was met the rules, and seven participants (with participants IDs: 3, 6, 7, 9, 17, 23, 30) were excluded. Therefore, in our study, the rest of the DEAP analysis continued with the remaining data of 25 participants.

2.4.2. SEED Dataset

SJTU Emotion EEG Dataset (SEED) is a public emotion dataset consist of fifteen subjects (7 males and eight females), aged 23.27±2.37 (mean ± std)) participating in the emotion-inducing experiment. Each subject was required to watch 15 selected film clips with positive, neutral, and negative stimuli to induce a corresponding emotional state with a duration of 4 minutes long of each film clip. Three sessions of data were collected, and each session comprises 15 trials/movies. The data were recorded using 62 EEG channels with a 1000 sampling Rate from each participant. The placements of 62 EEG channels were according to the international 10-20 system. Then, the data were downsampled to 200Hz to reduce computing complexity. A bandpass filter from 0-75Hz was applied.   In this paper, we only used the positive and negative labels/classes for a fair comparison of 2 class problems of the other emotional state datasets. In summary, 45 files (3 experiments for each participant) were used with data shape of (trail, channels, samples data) and label file of the 15 trails. More details about the dataset can be found in [45].

2.4.3. EDPMSC Dataset

The EEG Dataset for Classification of Perceived Mental Stress (EDPMSC) is a publicly available dataset that contains the EEG physiological signals of 28 participants (13 men and 15 women, ages 18–40) [19] . The EDPMSC contains data collected at 256 sampling rates from four Muse headband dry EEG channels. These electrodes are AF7, AF8, TP9, and TP10 and are placed on the scalp as references using Fz. The PSS questionnaire was used to assess a subject's level of stress over the preceding 30 days, which was then used to categorize EEG signals as stress or not stress. Each participant underwent three experiments. The first experiment was termed the pre-active phase, during which EEG data were collected for three minutes while sitting in a relaxed position in a quiet room with open eyes. The second phase involved recording EEG data during a presentation (activity phase) in front of a group of people. Finally, there was a three-minute post-activity phase during which EEG data were collected in the recording room. If the PSS score was greater than or equal to twenty, the subject was classified as stressed; if it was less than twenty, the subject was classified as non-stressed. The author of the dataset compared the pre-active and post-active phases and concluded that the pre-active phase is more accurate at identifying stress. As a result, we employ the precative phase in this study to develop our proposed model. The raw EEG data were preprocessed using a bandpass filter of a finite impulse response (FIR) filter with a bandwidth of 0.5 Hz to 35 Hz. To eliminate slow drifts, a high bandpass filter with a cutoff frequency of 0.5Hz was chosen. Additionally, the low bandpass filter of 35Hz was used to eliminate line noise at 50Hz and to store data for the delta, theta, alpha, and beta frequency bands.

Table 1 A SUMMARY DESCRIPTION OF THE DATASETS USED IN THIS STUDY

Dataset	Stimuli (stressor)	Stress labeling	Total EEG channels	Selected channels	No. Participants / Total Experiments	Frequency Rate (Hz)	classes
DEAP	Music video	SAM	32	AF3', 'FC5', 'F8', 'Fp1', 'AF4', 'P7', 'Fp2', 'F7	32/32	128	Stress/ Calm
SEED	Emotional video	Questionnaire	62	'AF3', 'FC5', 'F8', 'Fp1', 'AF4', 'P7', 'Fp2', 'F7'	15/45	200	Negative/Positive

| EDPMSC | History | PSS | 4 | 'TP9', 'AF7', 'AF8','TP10' | 28/84 | 256 | Stress/ not stress |
| Our | MA, Negative feedback and time pressure | Saliva cortisol | 7 | 'Fp1', 'Fp2', 'F7', 'F3', 'Fz', 'F4', 'F8' | 22/22 | 256 | Stress/ Rest |

3.   Methodology

A hybrid dominant feature selection method is developed in this study to enhance the classification performance of EEG mental stress recognition. A block diagram of the proposed method is shown in Figure 3. Multi-domain features are extracted from the time domain, frequency domain, time-frequency domain, and connectivity features. The optimal subset selected features are used to classify mental stress state using optimized SVM. The method's implementation phases are outlined below.

- Dataset Preprocessing
- Multi-domain features are extracted from multi-EEG channels and combined to form a large feature vector.
- Feature selection based on the proposed mRMR-PSO method is used to identify discriminative features.
- Classification parameters of SVM were optimized using PSO.
- The proposed model was validated with three different public datasets.

3.1. Data Preprocessing

The preprocessing of EEG signal was implemented using Python and an external package called MNE. Raw EEG signals were preprocessed using a bandpass filter. Finite impulse response (FIR) filters between 0.5 Hz to 35 Hz to remove DC artifacts and line noise (50/60Hz). All EEG channels were subjected to the common average reference. Fast, independent component analysis was used to eliminate the eye-blink generated by electrooculogram (EOG), eye movements as well as muscular artifacts. After that, the clean signals were employed for the rest of the research.

3.2. Feature Extraction

In this study, we extracted multi-domain features seeking a better presentation of the signal that provides acceptable accuracy for stress detection. Features from the time domain, frequency domain, time-frequency domain, and connectivity features were extracted. The connectivity features were estimated by utilizing Phase Locking Value. Features from the time-domain were based on Hjorth Parameters of activity mobility, and complexity, Peak to peak amplitude, Line Length, Kurtosis, and Skewness. Frequency domain features were based on the relative power of theta (4-8Hz), alpha(8-12Hz), sigma(12-15Hz), low beta(15-20Hz), and high beta (20-30Hz). Likewise, Time-Frequency domain features were based on Spectral Entropy (PSD, Welch) [12] and Katz Fractal Dimension [29]. Table 2 shows the summary of all features used in this study. These features were then normalized using feature-based z-score normalization to avoid large-scale weighting.

Table 2 Summary of Multi-Domain Feature Extraction Methods Employed in the Selected Datasets

Domain	Feature Name	Description	No. Features	Formula
Connectivity	Phase Locking Value [46]	It is a proportion of phase difference between signals over different trials above or below the 0 degree	$\frac{n(n-1)}{2}$	$PLV_{ij} = \left\| \frac{1}{T} \sum_{i=1}^{N} e^{i(\phi_i^a - \phi_j^a)} \right\|$

Time	Hjorth Parameters of activity mobility, and complexity [22]	Activity is the variance of the signal on time.	1	$Activity = \mathrm{var}(y(i))$
		Mobility represents the proportion of standard deviation of the window signal in the time domain.	1	$Mobility = \sqrt{\dfrac{\mathrm{var}'(y(i))}{Activity}}$
		Complexity represents how the shape of a signal is similar to a pure sine wave.	1	$Complexity = \sqrt{\dfrac{Mobility(\frac{dy(t)}{dt})}{Mobility(y(t))}}$
	Peak to peak amplitude	Represents the peak time of EEG signal between the various windows.	1	$PTP = pk_{high} - pk_{low}$
	Line Length [22,47]	Named a curve length, which indicates the total vertical length of the signal.	1	$L(n) = \sum_{i=1}^{N-1} x[i] - x[i-1]$
	Kurtosis [48]	Shows the sharpness of EEG signals' peaks.	1	$Kurtosis = \dfrac{\frac{1}{t}\sum_{t=1}^{T}(x(t)-\mu)^3}{\sigma 4}$
	Skewness	Represents the asymmetry of an EEG signal.	1	$Skewness = \dfrac{\frac{1}{t}\sum_{t=1}^{T}(x(t)-\mu)^3}{\sigma 3}$
Frequency	Relative Powers of [49]: theta (4-8Hz) alpha(8-12Hz) sigma(12-15Hz) low beta(15-20Hz) a high beta (20-30Hz).	Relative power represents the average absolute power of the given band intervals.	5	$RP = \dfrac{power(band)}{power(All\_bands)} * 100$
Time-Frequency	Spectral Entropy (PSD, Welch)[12]	Measure the distribution of signal power over frequency.	1	$SE = -\sum_{f=4}^{F=45} \overline{PSD}(F) log(\overline{PSD}(F))$
	Katz Fractal Dimension [29]	Represents the maximum distance between the first point and any other point of the Signal's time window.	1	$D = \dfrac{log_{10}(n)}{log_{10}(\frac{L}{d}) + log_{10}(n)}$

3.3. Feature Selection using mRMR-PSO

The proposed mRMR-PSO method consists of minimum redundancy maximum relevance (mRMR) [50] and a PSO algorithm. The proposed method aims to select the more informative feature subsets related to mental stress from the high dimensional space of EEG signal to improve the accuracy performance of the SVM classifier by ranking the relative and informative features. We utilize the filter method of mRMR first to generate a short feature pool and PSO wrapper to get the least redundant feature set and optimized SVM parameters for better accuracy. The sections below describe the details of each method.

3.3.1. minimum-Redundancy Maximum Relevance (mRMR)

The mRMR is a filter-based method, was first proposed by Ding and Peng,(2005)[51] as a solution for feature selection problems and shown to be computationally fast. The mRMR is used to rank a subset of features by minimizing the redundancy between the subset of features while maximizing the relevance of the feature to the target.

The proposed method employs mRMR to minimize the search space of the local optima of the original feature by ranking a subset of important features. The mRMR algorithm is based on a relevance measurement using an F-score, while the redundancy measurement is based on Pearson's correlation among features set [52]. This process minimizes the selection of redundant features, which results in minimizing the risk of overfitting at the PSO phase and solve the issue of global search space.

The maximum relevance (RL) is computed using F-statistic $F(X_i, y)$ between feature and target class   as the equation :

$$maxRL_f, RL_f = \frac{1}{|S|} \sum_{X_i \in S} F(X_i, y) \tag{3}$$

Where $S = \{X_1, X_2, X_3, ..., X_n\}$ is the set of features, y is the target class (ex. stress/ not. Stress), $|S|$ is the size of the feature set.

The minimum redundancy (RD) among features is computed using Pearson's correlation between pair of features as shown in equation(4):

$$\min RD_\rho, RD_\rho = \frac{1}{|S|^2} \sum_{X_i \in S} \rho(X_i, X_j) \tag{4}$$

The full join formula of the mRMR selection schema to rank the feature set is calculated using the F-test correlation quotient f^{FCQ} as the equation:

$$f^{FCQ}(X_i) = \max_{X_i \in \Omega_S} \left\{ \frac{F(X_i, y)}{\frac{1}{|S|} \sum_{X_j \in S} |\rho(X_i, X_j)|} \right\} \tag{5}$$

where $\rho(X_i, X_j)$ is the Pearson's correlation between pair of features, $F(X_i,y)$ is F-statistic, X_i (i ∈ {1, 2, ..., m}) is feature importance based on the mRMR criterion, m is total features, $|S|$ is the size of the feature set and y refer to target (class/label). In summary, at each stage of the mRMR feature selection process, the features with the highest feature important score will be added to the subset $f^{FCQ}(X_i)$ selected feature ranks. The f^{FCQ} mRMR feature selection results in achieving a more coverage balance in the solution space as well as contributing significant improvements to classification performance.

3.3.2. PSO Algorithm

Particle swarm optimization (PSO) was proposed by Kennedy & Eber-hart, (1995) for optimization problems [37,53–55]. PSO is a swarm intelligence meta-heuristic technique motivated by social behavior such as fish schooling, birds searching in search of food. PSO is based on the concept of birds exchanging information with one another. When birds seek food at random, they have no idea where to look. Like the evolutionary algorithm and genetic algorithm, PSO performs searches on a population (called swarm) of individuals (called particles), which are updated from iteration to iteration. PSO discovers the

optimal solution by allowing each particle changes its searching direction based on two factors, the best of all features (gbest) and its own best previous experience (pbest).

The status of each particle is characterized based on its position (global optima) and velocity (distance: local optima). If the position of each particle found its best position, then the information would be delivered to other particles. The particles' velocity and position were updated over iteration for the search of pbest and gbest for optimal solution p as equations.

$$V_{id}^{t+1} = w \cdot V_{id}^{t} + c_1 \cdot r_1(pbest_{id}^{t} - X_{id}^{t})$$
$$+ c_2 \cdot r_2(gbest_{id}^{t} - X_{id}^{t}) \tag{6}$$

$$X_{id}^{t+1} = X_{id}^{t} + V_{id}^{t+1}, d = \{1, 2, ..., D\} \tag{7}$$

where t denotes evolutionary generation, V_{id}^{t} *denotes particle's velocity i on dimension d*, X_{id}^{t} *denotes particle's position i on dimension d*, (c_1, c_2) denotes social learning factors of personal best (pbest) and global best (gbest) respectively, and (r_1, r_2) are random numbers of uniformly distributed U(0,1). The w refers to the weight used to balance global exploration and local exploitation.

3.3.3. Proposed Hybrid Method: mRMR-PSO-SVM.

In this section, we propose a hybrid method mRMR-PSO-SVM proposed for mental stress classification as shown in fig. The mRMR-PSO-SVM algorithm aims to select the optimal feature set from the reduced set of f^{FCQ} mRMR while optimize the classification performance by estimating the optimal values of SVM parameters (C, γ) simultaneously.

In our approach, three main phases are considered for better optimization: Initialization phase, Feature selection phase, and classification & evaluation phase.

In the initialization phase: it is proved that the high number of particles (P) increases the computational complexity of the optimization process while the small search space of P results in poor optimal solutions [55]. Therefore, we select 200 and 50 for a total number of generations (t) and a total number of particles, respectively. Likewise, (c_1, c_2) and w values influence the convergence of the optimization process. If they are set too high, the particle velocity becomes too fast, and the optimum solution cannot be obtained. Thus, we set (c_1, and c_2) to 2 and w set to mean the mutual information of the subset selection f^{FCQ}.

In the feature selection phase:

The mRMR algorithm is developed as described in Section B.1 to rank the most important features of the given datasets that result in the highest classification performance with the SVM classifier. The mRMR-PSO evaluate each selection of ranked feature subset and SVM parameters according to a fitness function which is the classification F-measure of SVM.

Classification and evaluation phase:

support vector machine (SVM) has been widely used in different applications, including EEG-based applications [56,57]. The classification procedure is a part of wrapper feature selection methods to evaluate and validate the model. In our method, we optimize SVM parameters (C, γ) using PSO and evaluate the subset performance using the activation function of F-measure as the equations (8-10):
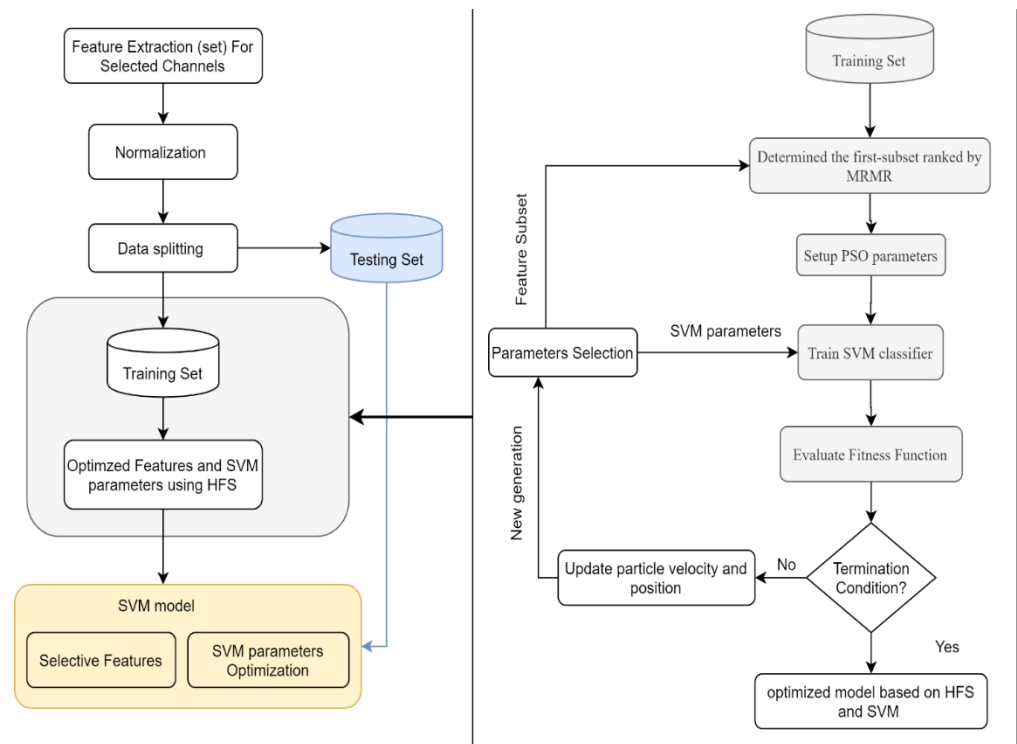
Figure 3 The flowchart of the proposed feature selection method' mRMR-PSO-SVM.

Algorithm 1: An Improved Hybrid Feature Selection Algorithm based on mRMR-PSO-SVM

Result: Best feature Subset(U)

Input: N: multi-domain feature set of selected channels and and label (model 1)

Initialization: PSO parameters (populationsMax =200 , particlesMax =50, learning parameters (C_1, C_2) = 2, the inertia weight (w) = mean subset of mRMR and random generate velocity and position of each particle);

while *i:populationsMax or stopping criterion is not met* **do**

⎢   ●Step 1: mRMR method to rank subset features of N equation (1).
⎢   ●Step 2: Train the selected subset of the features from step 1 using $RBF - SVM$.
⎢   ●Step 3: Evaluate the selected features with PSO and SVM parameters (C, y) equation (2).
⎢   ●Step 4: Update PSO (velocity and position) and SVM (C,y) using the equation(3,4).
⎢   ●Step 6: termination condition: recursively use steps 1-4 till the criteria of termination are satisfied (e.x number of generations or accuracy fulfilled).

end

$$precision = \frac{TP}{TP + FP} \qquad (8)$$

$$recall = \frac{TP}{TP + FN} \tag{9}$$

$$F - measure = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{10}$$

Where TP refers to the total number of true positives, false positive (FP), true negative (TN), and false-negative (FN).

Accuracy is also used as an overall measure for classification, which is:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

Additionally, SVM has different kernel functions such as linear, polynomial, and radial basis functions (RBF). In this study, we utilize RBF to obtain optimal solution due its widely used when dealing with multi-dimensionality space, also the number of the parameter that needs to be optimized are few compared to other kernels such as polynomial. RBF has two parameters C and γ, Parameter C denotes the cost of the penalty. The choice of value for C influences the classification outcome. While Parameter γ has a much greater influence on classification outcomes than C because its value affects the partitioning outcome in the feature space [58]. The primary aim of selecting a suitable kernel function along with its kernel parameter (s), because the kernel defines the feature space in which the training sets will be classified. In this context, the values of RBF parameters need to be optimized for the optimal use of the SVM along with the feature selection.

In summary, the main basic procedure for the proposed algorithm (mRMR-PSO-SVM) is presented as follows:

Step 1: Use the mRMR method to rank the features of the training sets from the highest best feature to the lowest using the equation (**5**).

Step 2: Initialize PSO parameters (populations, number of particles, learning parameters (c_1, and c_2), the inertia weight (w) and the generate velocity and position of each particle).

Step 3: Train the selected subset of the features from step 1 using RBF-SVM.

Step 4: Evaluate the selected features with PSO and SVM parameters using the fitness function of F-measure as shown in equation (10).

Step 5: Update parameters of PSO (velocity and position) and SVM parameters (C, γ) till the termination criteria are met.

Step 6: termination condition: recursively use steps 3,4 and 5 for refining the model fitness till the criteria of termination are satisfied (e.x number of generations or accuracy fulfilled).

Step 7: Classify mental stress from testing data using generated optimal model (optimal selected model and optimized parameters).

4. Result

In this section, we present the statistical results of the mental stress experiment induced by MAT and assessed by EEG and alpha-amylase. Although, we report the mental stress state classification based on optimal feature set, selected from multi-domain

features, of network connectivity features, time domain, frequency domain, and time-fre-quency domains, using the proposed method mRMR-PSO-SVM.

4.1. Statistical Analysis

In our EDMSS experiment, the stress tasks were induced using a mental arithmetic task with negative feedback and time pressure. The salivary alpha-amylase (SAA) was used to assess and validate mental stress during EEG acquisition and plays as a bi-omarker for EEG annotation. The mean scores acquired from 22 participants using the SAA are shown in Figure 4.
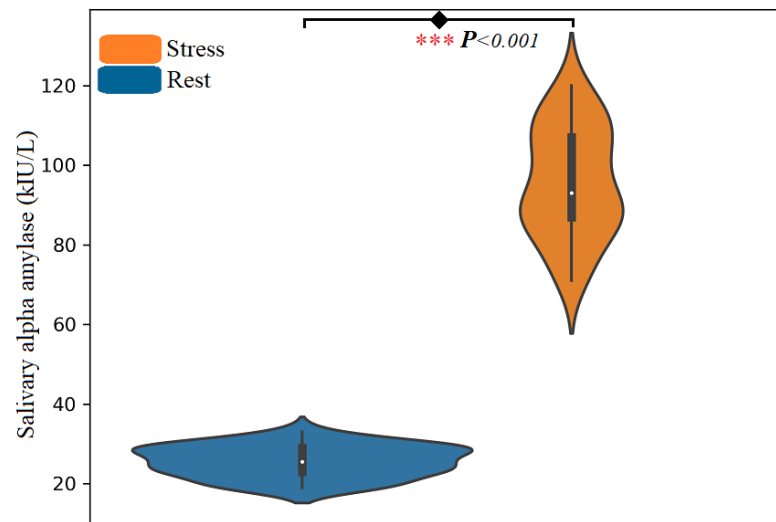
Figure 4. The average score of salivary alpha-amylase level responses for stress and rest tasks. Two measurement samples (5 min before (baseline), and 5 min after the last stress task. The "*******" marks indicate, the task is significant with p<0.001.

Overall, the reported SAA level among participants scores $(\mu, \sigma) = (24.45 \pm 4.44$ (kIU/L)) before stress inducement and $(\mu, \sigma) = (93.64 \pm 13.99$ (kIU/L)) after stress inducements. Participants with an SAA score of more than 60 (kIU/L) were classified as being in a stress state, whereas those with a score of less than 30 (kIU/L) were classified as being in the rest group. The T-test was applied on SAA to verify the affective of stress inducements on rest and stress states. The difference between the mental states is considered significantly different if the p-value is less than 0.05. The results revealed a significant difference between the two states with p<0.001. The approach of stress inducement using mental arithmetic tasks used in this experiment is similar to that used by other researchers [16].

4.2. Performance analysis of feature selection and multi-domain features

We evaluate the performance of our methodology using EDMSS and further validated using three public EEG datasets. The datasets DEAP, SEED, and EDPMSC were utilized here for mental stress recognition. A summary of the datasets is provided in The EEG Dataset for Classification of Perceived Mental Stress (EDPMSC) is a publicly available dataset that contains the EEG physiological signals of 28 participants (13 men and 15 women, ages 18–40) [19] . The EDPMSC contains data collected at 256 sampling rates from four Muse headband dry EEG channels. These electrodes are AF7, AF8, TP9, and TP10 and are placed on the scalp as references using Fz. The PSS questionnaire was used to assess a subject's level of stress over the preceding 30 days, which was then used to categorize EEG signals as stress or not stress. Each participant underwent three experiments. The first experiment was termed the pre-active phase, during which EEG data were collected for three minutes while sitting in a relaxed position in a quiet room with open eyes. The second phase

involved recording EEG data during a presentation (activity phase) in front of a group of people. Finally, there was a three-minute post-activity phase during which EEG data were collected in the recording room. If the PSS score was greater than or equal to twenty, the subject was classified as stressed; if it was less than twenty, the subject was classified as non-stressed. The author of the dataset compared the pre-active and post-active phases and concluded that the pre-active phase is more accurate at identifying stress. As a result, we employ the precative phase in this study to develop our proposed model. The raw EEG data were preprocessed using a bandpass filter of a finite impulse response (FIR) filter with a bandwidth of 0.5 Hz to 35 Hz. To eliminate slow drifts, a high bandpass filter with a cutoff frequency of 0.5Hz was chosen. Additionally, the low bandpass filter of 35Hz was used to eliminate line noise at 50Hz and to store data for the delta, theta, alpha, and beta frequency bands.

Table 1 shows the number of EEG channels used from each dataset. In DEAP and SEED datasets eight channels were selected mostly from the prefrontal and frontal regions of the brain, seven EEG channels from EDMSS, and 4 EEG channels from EDPMSC.

The multi-domain features were extracted from each dataset and used as input vectors after normalizing them using column-base z-score normalization. The features of multi-domains are combined to derive a high-dimensional feature vector.

Table 2 shows the summary of the proposed multi-domain features presenting the domain name, feature description, total number features, and feature formula is provided. A total of 161 multi-domain features were extracted from the seven EEG channels of EDMSS, 188 features for the datasets containing 8 channels (DEAP, SEED), and only 86 features were extracted from the EDPMSC dataset containing 4 EEG channels.

Figure 5 and Figure 6 represent the results of mRMR-PSO-SVM in selecting the optimal feature subset per dataset. From the figures, it can be observed that mRMR-PSO-SVM can significantly reduce a large number of features vector space while achieving a high classification performance. Figure 6 shows the classification performance results, and the number of selected features obtained by our proposed algorithm on different datasets. The highest average classification performance on EDMS was 77.23%, 80.87%, 76.30%, and 77.41% for accuracy, precision, recall, and f1-score respectively with an average of 52 optimal selected features. Whereas, in the DEAP dataset, the proposed algorithm achieved an average performance of 93.88%, 91.11%, 94.91 %, and 91.99% for accuracy, precision, recall, and f1-score respectively with an average of 56 selected features. Additionally, for the SEED dataset, the achieved results were 84.17 for accuracy, 90% precision, 83.23% recall, and 85.54 f1-score using 49 optimal selected features. In the EDPMSC dataset, the results obtained, based on dependent subject analysis since each subject was labeled as a stressed subject or not, were 89.31%, 85.11%, 85.11 % 85.11% for accuracy, precision, recall, and f1-score respectively with only 45 selected features.

From Figure 5 it can be observed that the proposed algorithm reduced the total number of features to less than 70% from the original feature vector space of all datasets while increasing the prediction accuracy or maintain it. On all four datasets, the mRMR-PSO-SVM approach preserved just around 30% of the features while improving prediction accuracy. The findings indicate that the proposed mRMR-PSO-SVM method is capable of efficiently removing redundant or irrelevant features, resulting in better classification performance.
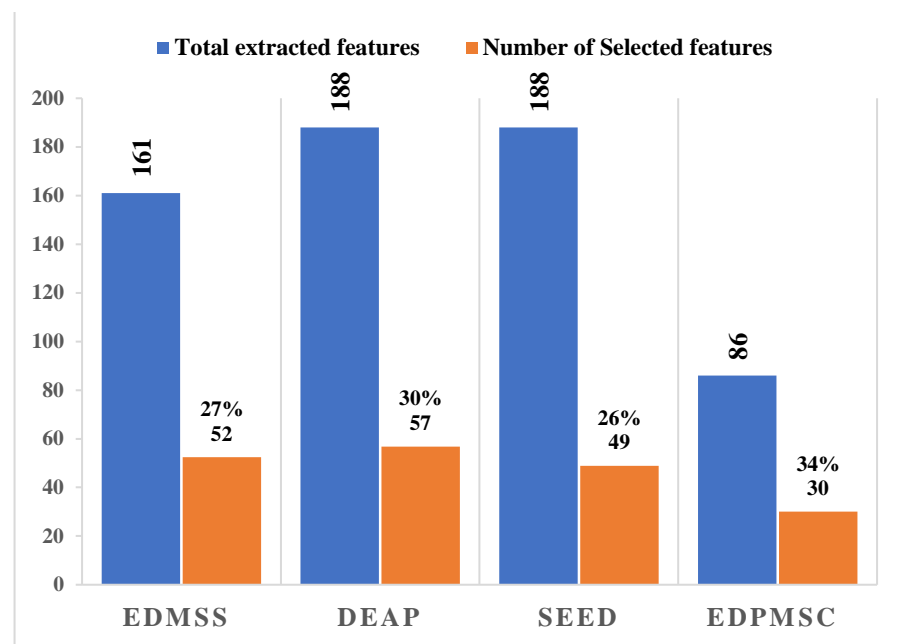
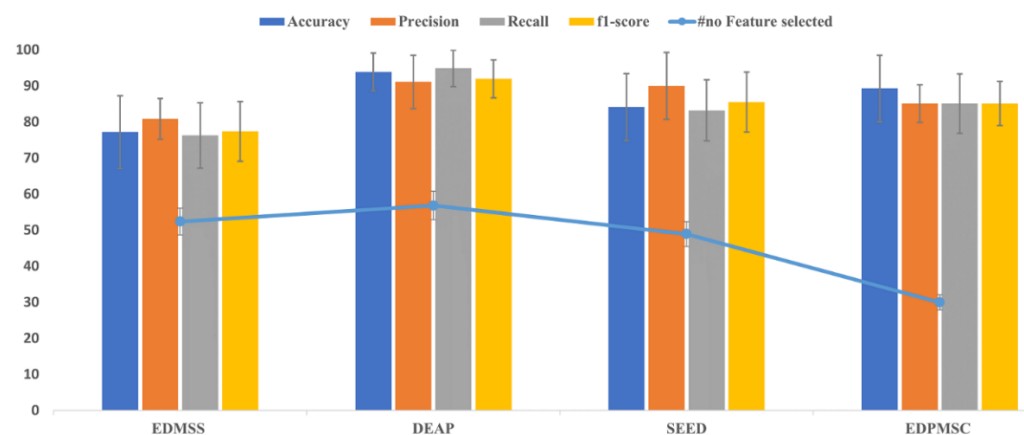Figure 5 A total number of multi-domain features selected using mRMR-PSO-SVM

Figure 6 The evaluation performance of the proposed mRMR-PSO-SVM on different EEG datasets for mental stress detection.

To evaluate the effectiveness of the proposed mRMR-PSO-SVM method was compared to other state-of-the-art swarm intelligence meta-heuristic algorithms, namely, BAT, FFA, GWO, MFO, MVO, PSO, and WOA [59]. We assess the performance of each feature selection method using three parameters: classification accuracy, execution time, and the number of selected features utilized for mental stress classification. For that, a split mechanism was used to train and test feature selection methods for obtaining the classification accuracy with 80% for training and 20% for testing.

Table 3 shows the average number of the accuracy, selected features, and execution time for each swarm optimizer algorithms with EDMSS, DEAP, SEED, and EDPMSC datasets. In DEAP, the proposed algorithm achieved the highest accuracy of 93.878% using 57 selected features while the highest among the compared algorithms was MVO with 88.877% accuracy and 86 features. Meanwhile, in the SEED dataset, the proposed method

obtained an accuracy of 84.167% with only 49 selected features which are higher than the best competitive optimizer FFA that performed 74.815% of accuracy using 90 features with the same dataset. Additionally, mRMR-PSO-SVM obtained 77.22%, 88.301% accuracies with 52 and 30 selected features for both EDMSS and EDPMSC, respectively. However, in EDPMSC, the WOA showed a slight increase in accuracy with 1.026% using 36 selected features.

Table 3 The average values of the statistical parameters of classifiers using the subject independent test.

Algorithm	Execution time	Accuracy	#No Selected features	Execution time	Accuracy	#No Selected features
	EDMSS DATASET			**EDPMSC DATASET**		
BAT	4.315	67.624	75	15.378	87.703	44
FFA	19.615	65.172	79	19.285	87.935	36
GWO	9.234	67.664	74	15.001	87.703	55
MFO	4.336	67.267	85	16.586	88.167	55
MVO	4.135	67.631	80	14.620	88.863	45
PSO	5.530	65.289	108	15.923	84.919	55
WOA	5.773	64.224	72	15.195	89.327	36
Proposed	**11.719**	**77.222**	**52**	**60.700**	**88.301**	**30**
	DEAP DATASET			**SEED DATASET**		
BAT	10.328	88.229	80	2.946	68.889	86
FFA	41.391	88.079	87	14.852	74.815	90
GWO	21.013	87.515	83	6.939	71.111	84
MFO	46.348	88.182	97	2.865	70.370	85
MVO	10.695	88.877	86	2.869	70.370	85
PSO	13.682	88.276	121	4.027	66.667	122
WOA	14.482	88.697	79	4.236	68.148	79
Proposed	**53.768**	**93.878**	**57**	**9.346**	**84.167**	**49**

5. Discussion

The main object of mRMR-PSO-SVM is to select an optimal number of features that increase or maintain the overall performance by enhancing the exploration, exploitation of search space.

Various time-domain, frequency-domain, and time-frequency-domain features have been proposed in previous studies for mental stress detection [13,15,16]. However, little research utilizes connectivity features. In our study, we utilize the most important features from multiple domains, seeking better informative features for stress detection. As a result, a fusion of multi-domain features showed a promising result in different fields as there could be multi-way interactions among features [14,60]. The drawback of multi-domain features is that they are vulnerable to redundant and unrelated features. Therefore, the proposed mRMR-PSO-SVM method aims to select the discriminative features from high dimensional space. The optimal selected features using mRMR-PSO are used for training SVM with RBF kernel, followed by the evaluation of the classification performance with test data. In each dataset, an independent subject test was conducted.

Table 3 shows that the proposed mRMR-PSO-SVM outperforms the other algorithms, used in this study, in terms of accuracy and selected features. The significant advantage in finding the optimum number of features is justified since the proposed mRMR-

PSO-SVM employs the strength of mRMR and PSO mechanisms, which enhance the possibility of selecting weak solutions and decrease the probability of trapping in local optima. This method allows the algorithm to fully explore parts of the feature space before using PSO to improve these regions.

To address feature selection problems, this approach makes use of the strengths of a global search algorithm, which is efficient in both exploration (local search) and exploitation (global search). Finding fewer optimum features means that the algorithm has successfully removed irrelevant/redundant features from the feature vector space of the dataset. However, for excursion time it shows slightly better than FFA in EDMSS and SEED dataset while it takes more time compared to other optimizers. This could be due to the hybrid interactions between the filter method (mRMR) and wrapper method (PSO).

It is worth mentioning that the proposed mRMR-PSO method outperforms the conventional PSO algorithm in classification accuracy as well as most the state-of-the-art meta-heuristic algorithms, namely, BAT, FFA, GWO, MFO, MVO, PSO, and WOA [59]. As a result, the mRMR-PSO is a promising technique for detecting significant factors while removing redundant and irrelevant data. When compared to the original PSO algorithm with the same approach, we can conclude that the proposed mRMR-PSO method offers accurate classification performance with the fewest number of selected features in all datasets. Thus, the mRMR-PSO algorithm provides a viable solution to feature selection and mental stress classification problems.

To further demonstrate the success of the proposed hybrid feature selection method, the mRMR-PSO-SVM showed superior performance in comparison to state-of-the-art feature selection methods in terms of accuracy, number of selected features as listed in Table 4. From Table 3 and Table 4, we can highlight that our proposed method, outperform other proposed methods in term of accuracy and number of selected features. This could be due to many factors tacked in our model such as fusion of multi-domain features, improving the exploitation of PSO, and optimizing the SVM parameters.

Table 4 Comparison with previous studies on related public available datasets for mental stress detection.

#Ref.	Dataset	FS- Classifier	Total feature vector / Selected Features	No. Channels	Accuracy
[44]	DEAP	GA- KNN[1]	673/not mentioned	32	71.76%
[17]	DEAP	Boruta-KNN	608/288*	32	73.38%
[19]	EDPMSC	Wrapper FS- (MLP, SVM)	90/18	4	89.30% MLP, 67.85% SVM for pre-active phase
[61]	DEAP	2-D AlexNet-CNN 3-D AlexNet-CNN[2]	5 PSD bands converted to image	32	84.77%, 86.12%
[62]	SEED, DEAP	DWT-BODF[3] (SVM, KNN)	225 × 30 SEED 576 × 40 DEAP	62 SEED 32 DEAP	93.8% SVM (SEED) 77.4% SVM (DEAP)

It's worth emphasizing that our proposed model focused on EEG based mental stress only, however, it could be further utilized with other datasets from different domains.

To summarize, the study's findings are quite promising. However, there is still much potential for additional research in the field of EEG-based mental stress classifications. To begin, we conducted our study using a fixed time window length; however, experimenting with alternative window lengths can aid in model generalization. Additionally, It is worth exploring, data augmentation via sliding windows as this technique is increasingly being utilized in combination with deep learning on EEG [61]. Second, a more in-depth

examination of other connective network features should be explored since they give detailed information about channel interconnections. Our method was based on established signal noise removal techniques [16]. However, alternative denoising techniques such as signal smoothing should be considered. Additionally, prior research has demonstrated that utilizing feature extraction with feature selection approaches and neural networks results in high accuracy in EEG based models. As a result, future studies may potentially explore neural networks and deep learning approaches.

3. Conclusions

In this paper, a hybrid feature selection method, mRMR-PSO-SVM, was proposed to select the most informative features related to the mental stress task. By ranking important features as a subset of the original feature set, the mRMR was used to reduce the search space of the local optima of the original feature set. Following that, -PSO evaluates each ranked feature subset selection and optimizes SVM parameters according to a fitness function. On the basis of selected features and classification performance, the mRMR-PSO-SVM was evaluated using four datasets and compared to existing methods. The proposed method outperformed other feature selection methods in terms of accuracy and number of selected channels, according to the experimental results obtained on various datasets. Despite the fact that the proposed method yielded promising results, future research could further validate the methods with different domain datasets and use neural networks and deep learning approaches to better evaluate the quality of selected features and their effects on computation time.

References

1. Rostami, M.; Berahmand, K.; Nasiri, E.; Forouzande, S. Review of swarm intelligence-based feature selection methods. *Eng. Appl. Artif. Intell.* **2021**, *100*, doi:10.1016/j.engappai.2021.104210.

2. Hussien, A.G.; Oliva, D.; Houssein, E.H.; Juan, A.A.; Yu, X. Binary whale optimization algorithm for dimensionality reduction. *Mathematics* **2020**, *8*, 1–24, doi:10.3390/math8101821.

3. Kang, X.; Handayani, D.O.D.; Chong, P.P.; Acharya, U.R. Profiling of pornography addiction among children using EEG signals: A systematic literature review. *Comput. Biol. Med.* **2020**, *125*, 103970, doi:10.1016/J.COMPBIOMED.2020.103970.

4. Pei, Z.; Wang, H.; Bezerianos, A.; Li, J. EEG-Based Multiclass Workload Identification Using Feature Fusion and Selection. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 0–9, doi:10.1109/TIM.2020.3019849.

5. Tuncer, T.; Dogan, S.; Subasi, A. EEG-based driving fatigue detection using multilevel feature extraction and iterative hybrid feature selection. *Biomed. Signal Process. Control* **2021**, *68*, 102591, doi:10.1016/j.bspc.2021.102591.

6. Angga Yuwono, H.; Kusuma Wijaya, S.; Prajitno, P. Feature selection with Lasso for classification of ischemic strokes based on EEG signals. *J. Phys. Conf. Ser.* **2020**, *1528*, doi:10.1088/1742-6596/1528/1/012029.

7. Molla, M.K.I.; Shiam, A. Al; Islam, M.R.; Tanaka, T.; Tanaka, T.; Tanaka, T. Discriminative Feature Selection-Based Motor Imagery Classification Using EEG Signal. *IEEE Access* **2020**, *8*, 98255–98265, doi:10.1109/ACCESS.2020.2996685.

8. Smitha, K.G.; Vinod, A.P.; Mahesh K Voice familiarity detection using EEG-based Brain-Computer Interface. In Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC); IEEE, 2016; pp. 001626–001631.

9. Tzimourta, K.D.; Astrakas, L.G.; Gianni, A.M.; Tzallas, A.T.; Giannakeas, N.; Paliokas, I.; Tsalikakis, D.G.; Tsipouras, M.G. Evaluation of window size in classification of epileptic short-term EEG signals using a Brain Computer Interface software. *Eng. Technol. Appl. Sci. Res.* **2018**, *8*, 3093–3097, doi:10.48084/etasr.2031.

10. Sun, Z.; Huang, Z.; Duan, F.; Liu, Y. A Novel Multimodal Approach for Hybrid Brain–Computer Interface. *IEEE Access* **2020**, *8*, 89909–89918, doi:10.1109/ACCESS.2020.2994226.

11. Movahed, R.A.; Jahromi, G.P.; Shahyad, S.; Meftahi, G.H. A major depressive disorder classification framework based on EEG signals using statistical, spectral, wavelet, functional connectivity, and nonlinear analysis. *J. Neurosci. Methods* **2021**, *358*, 109209, doi:10.1016/j.jneumeth.2021.109209.

12. Yin, Y.; Zheng, X.; Hu, B.; Zhang, Y.; Cui, X. EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM. *Appl. Soft Comput.* **2021**, *100*, 106954, doi:10.1016/j.asoc.2020.106954.

13. Halim, Z.; Rehan, M. On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning. *Inf. Fusion* **2020**, *53*, 66–79, doi:10.1016/j.inffus.2019.06.006.

14. Hag, A.; Handayani, D.; Pillai, T.; Mantoro, T.; Kit, M.H.; Al-Shargie, F. EEG Mental Stress Assessment Using Hybrid Multi-Domain Feature Sets of Functional Connectivity Network and Time-Frequency Features. *Sensors* **2021**, *21*, 6300, doi:10.3390/s21186300.

15. Martínez-Rodrigo, A.; García-Martínez, B.; Huerta, Á.; Alcaraz, R. Detection of negative stress through spectral features of electroencephalographic recordings and a convolutional neural network. *Sensors* **2021**, *21*, doi:10.3390/s21093050.

16. Katmah, R.; Al-Shargie, F.; Tariq, U.; Babiloni, F.; Al-Mughairbi, F.; Al-Nashash, H. A Review on Mental Stress Assessment Methods Using EEG Signals. *Sensors* **2021**, *21*, 5043, doi:10.3390/s21155043.

17. Hasan, M.J.; Kim, J.M. A hybrid feature pool-based emotional stress state detection algorithm using EEG signals. *Brain Sci.* **2019**, *9*, doi:10.3390/brainsci9120376.

18. Subhani, A.R.; Mumtaz, W.; Saad, M.N.B.M.; Kamel, N.; Malik, A.S. Machine Learning Framework for the Detection of Mental Stress at Multiple Levels. *IEEE Access* **2017**, *5*, 13545–13556, doi:10.1109/ACCESS.2017.2723622.

19. Arsalan, A.; Majid, M.; Butt, A.R.; Anwar, S.M. Classification of Perceived Mental Stress Using A Commercially Available EEG Headband. *IEEE J. Biomed. Heal. Informatics* **2019**, *23*, 2257–2264, doi:10.1109/JBHI.2019.2926407.

20. Al-Shargie, F.; Tang, T.B.; Kiguchi, M. Stress Assessment Based on Decision Fusion of EEG and fNIRS Signals. *IEEE Access* **2017**, *5*, 19889–19896, doi:10.1109/ACCESS.2017.2754325.

21. Cvetkovic, D.; Übeyli, E.D.; Cosic, I. Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: A pilot study. *Digit. Signal Process. A Rev. J.* **2008**, doi:10.1016/j.dsp.2007.05.009.

22. Boonyakitanont, P.; Lek-uthai, A.; Chomtho, K.; Songsiri, J. A review of feature extraction and performance evaluation in

epileptic seizure detection using EEG. *Biomed. Signal Process. Control* **2020**, *57*, 101702, doi:10.1016/j.bspc.2019.101702.

23.   Toradmalle, D.; Muthukuru, J.; Sathyanarayana, B. Hybrid Feature Selection Method based on Particle Swarm Optimization and Adaptive local Search Method. *Int. J. Electr. Comput. Eng.* **2019**, *9*, 3228–3231, doi:: 10.11591/ijece.v11i3.pp2414-2422.

24.   Mafarja, M.M.; Mirjalili, S. Hybrid Whale Optimization Algorithm with simulated annealing for feature selection. *Neurocomputing* **2017**, *260*, 302–312, doi:10.1016/j.neucom.2017.04.053.

25.   Ghanem, W.A.H.M.; Jantan, A. Novel multi-objective artificial bee colony optimization for wrapper based feature selection in intruction detectoin. *Int. J. Adv. Soft Comput. its Appl.* **2016**, *8*, 70–81.

26.   Xue, B.; Zhang, M.; Member, S.; Browne, W.N. Particle Swarm Optimization for Feature Selection in Classification : A Multi-Objective Approach. *Ieee Trans. Cybern.* **2012**, 1–16.

27.   Venkatesh, B.; Anuradha, J. A review of Feature Selection and its methods. *Cybern. Inf. Technol.* **2019**, *19*, 3–26, doi:10.2478/CAIT-2019-0001.

28.   Ma, W.; Zhou, X.; Zhu, H.; Li, L.; Jiao, L. A two-stage hybrid ant colony optimization for high-dimensional feature selection. *Pattern Recognit.* **2021**, *116*, 107933, doi:10.1016/j.patcog.2021.107933.

29.   Garro, B.A.; Salazar-Varas, R.; Vazquez, R.A. EEG Channel Selection using Fractal Dimension and Artificial Bee Colony Algorithm. *Proc. 2018 IEEE Symp. Ser. Comput. Intell. SSCI 2018* **2019**, 499–504, doi:10.1109/SSCI.2018.8628837.

30.   Mirjalili, S. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowledge-Based Syst.* **2015**, *89*, 228–249, doi:10.1016/j.knosys.2015.07.006.

31.   Jangir, P.; Parmar, S.A.; Trivedi, I.N.; Bhesdadiya, R.H. A novel hybrid Particle Swarm Optimizer with multi verse optimizer for global numerical optimization and Optimal Reactive Power Dispatch problem. *Eng. Sci. Technol. an Int. J.* **2017**, *20*, 570–586, doi:10.1016/j.jestch.2016.10.007.

32.   Sadeghian, Z.; Akbari, E.; Nematzadeh, H. A hybrid feature selection method based on information theory and binary butterfly optimization algorithm. *Eng. Appl. Artif. Intell.* **2021**, *97*, 104079, doi:10.1016/j.engappai.2020.104079.

33.   Bablani, A.; Edla, D.R.; Tripathi, D.; Dodia, S.; Chintala, S. A Synergistic Concealed Information Test with Novel Approach for EEG Channel Selection and SVM Parameter Optimization. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 3057–3068, doi:10.1109/TIFS.2019.2913798.

34.   Naserbegi, A.; Aghaie, M.; Zolfaghari, A. Implementation of Grey Wolf Optimization (GWO) algorithm to multi-objective loading pattern optimization of a PWR reactor. *Ann. Nucl. Energy* **2020**, *148*, 107703, doi:10.1016/j.anucene.2020.107703.

35.   Ahmed, M.A.; Qi, D.; Alshemmary, E.N. Effective hybrid method for the detection and rejection of electrooculogram (EOG) and power line noise artefacts from electroencephalogram (EEG) mixtures. *IEEE Access* **2020**, *8*, 202919–202932, doi:10.1109/ACCESS.2020.3036134.

36.   Agrawal, P.; Abutarboush, H.F.; Ganesh, T.; Mohamed, A.W. Metaheuristic algorithms on feature selection: A survey of one decade of research (2009-2019). *IEEE Access* **2021**, *9*, 26766–26791, doi:10.1109/ACCESS.2021.3056407.

37.   Ji, B.; Lu, X.; Sun, G.; Zhang, W.; Li, J.; Xiao, Y. Bio-Inspired Feature Selection: An Improved Binary Particle Swarm Optimization Approach. *IEEE Access* **2020**, *8*, 85989–86002, doi:10.1109/ACCESS.2020.2992752.

38.   Al-shargie, F.; Tang, T.B.; Badruddin, N.; Kiguchi, M. Towards multilevel mental stress assessment using SVM with ECOC : an EEG approach. *Med. Biol. Eng. Comput.* **2018**, 125–136, doi:10.1007/s11517-017-1733-8.

39.   Al-Shargie, F.; Tang, T.B.; Badruddin, N.; Kiguchi, M. Simultaneous measurement of EEG-fNIRS in classifying and localizing brain activation to mental stress. In Proceedings of the IEEE 2015 International Conference on Signal and Image Processing Applications, ICSIPA 2015 - Proceedings; 2016; pp. 282–286.

40.   Al-Shargie, F.; Tang, T.B.; Kiguchi, M. Assessment of mental stress effects on prefrontal cortical activities using canonical correlation analysis: an fNIRS-EEG study. *Biomed. Opt. Express* **2017**, *8*, 2583, doi:10.1364/BOE.8.002583.

41.   Al-Shargie, F.; Kiguchi, M.; Badruddin, N.; Dass, S.C.; Hani, A.F.M.; Tang, T.B. Mental stress assessment using simultaneous

measurement of EEG and fNIRS. *Biomed. Opt. Express* **2016**, *7*, 3882, doi:10.1364/BOE.7.003882.

42. Koelstra, S.; Muhl, C.; Soleymani, M.; Jong-Seok Lee; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. DEAP: A Database for Emotion Analysis ;Using Physiological Signals. *IEEE Trans. Affect. Comput.* **2012**, *3*, 18–31, doi:10.1109/T-AFFC.2011.15.

43. Özerdem, M.S.; Polat, H. Emotion recognition based on EEG features in movie clips with channel selection. *Brain Informatics* **2017**, *4*, 241–252, doi:10.1007/s40708-017-0069-3.

44. Shon, D.; Im, K.; Park, J.H.; Lim, D.S.; Jang, B.; Kim, J.M. Emotional Stress State Detection Using Genetic Algorithm-Based Feature Selection on EEG Signals. *Int. J. Environ. Res. Public Health* **2018**, *15*, doi:10.3390/ijerph15112461.

45. Zheng, W.L.; Zhu, J.Y.; Lu, B.L. Identifying stable patterns over time for emotion recognition from eeg. *IEEE Trans. Affect. Comput.* **2019**, *10*, 417–429, doi:10.1109/TAFFC.2017.2712143.

46. Al-Shargie, F.; Tariq, U.; Alex, M.; Mir, H.; Al-Nashash, H. Emotion Recognition Based on Fusion of Local Cortical Activations and Dynamic Functional Networks Connectivity: An EEG Study. *IEEE Access* **2019**, *7*, 143550–143562, doi:10.1109/ACCESS.2019.2944008.

47. Esteller, R.; Echauz, J.; Tcheng, T.; Litt, B.; Pless, B. Line length: An efficient feature for seizure onset detection. *Annu. Int. Conf. IEEE Eng. Med. Biol.* **2001**, *2*, 1707–1710, doi:10.1109/IEMBS.2001.1020545.

48. Alimardani, F.; Cho, J.H.; Boostani, R.; Hwang, H.J. Classification of bipolar disorder and schizophrenia using steady-state visual evoked potential based features. *IEEE Access* **2018**, *6*, 40379–40388, doi:10.1109/ACCESS.2018.2854555.

49. Asif, A.; Majid, M.; Anwar, S.M. Human stress classification using EEG signals in response to music tracks. *Comput. Biol. Med.* **2019**, *107*, 182–196, doi:10.1016/j.compbiomed.2019.02.015.

50. Direito, B.; Duarte, J.; Teixeira, C.; Schelter, B.; Le Van Quyen, M.; Schulze-Bonhage, A.; Sales, F.; Dourado, A. Feature selection in high dimensional EEG features spaces for epileptic seizure prediction. *IFAC Proc. Vol.* **2011**, *44*, 6206–6211, doi:10.3182/20110828-6-IT-1002.03331.

51. Hanchuan Peng; Fuhui Long; Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238, doi:10.1109/TPAMI.2005.159.

52. Zhao, Z.; Anand, R.; Wang, M. Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform. *Proc. - 2019 IEEE Int. Conf. Data Sci. Adv. Anal. DSAA 2019* **2019**, 442–452, doi:10.1109/DSAA.2019.00059.

53. Mehmood, R.M.; Lee, H.J. Emotion recognition from EEG brain signals based on particle swarm optimization and genetic search. *2016 IEEE Int. Conf. Multimed. Expo Work. ICMEW 2016* **2016**, doi:10.1109/ICMEW.2016.7574682.

54. Wu, S.L.; Liu, Y.T.; Hsieh, T.Y.; Lin, Y.Y.; Chen, C.Y.; Chuang, C.H.; Lin, C.T. Fuzzy Integral with Particle Swarm Optimization for a Motor-Imagery-Based Brain-Computer Interface. *IEEE Trans. Fuzzy Syst.* **2017**, *25*, 21–28, doi:10.1109/TFUZZ.2016.2598362.

55. Yang, S.T.; Lee, J. Der; Chang, T.C.; Huang, C.H.; Wang, J.J.; Hsu, W.C.; Chan, H.L.; Wai, Y.Y.; Li, K.Y. Discrimination between Alzheimer's disease and mild cognitive impairment using SOM and PSO-SVM. *Comput. Math. Methods Med.* **2013**, *2013*, doi:10.1155/2013/253670.

56. Chen, G.; Chen, J. A novel wrapper method for feature selection and its applications. *Neurocomputing* **2015**, *159*, 219–226, doi:10.1016/j.neucom.2015.01.070.

57. Li, X.; Hu, B.; Sun, S.; Cai, H. EEG-based mild depressive detection using feature selection methods and classifiers. *Comput. Methods Programs Biomed.* **2016**, *136*, 151–161, doi:10.1016/j.cmpb.2016.08.010.

58. Lin, S.W.; Ying, K.C.; Chen, S.C.; Lee, Z.J. Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Syst. Appl.* **2008**, *35*, 1817–1824, doi:10.1016/j.eswa.2007.08.088.

59. Khurma, R.A.; Aljarah, I.; Sharieh, A.; Mirjalili, S. EvoloPy-FS: An Open-Source Nature-Inspired Optimization Framework

in Python for Feature Selection. In; 2020; pp. 131–173.

60.   Tong, L.; Zhao, J.; Fu, W. Emotion Recognition and Channel Selection Based on EEG Signal. *Proc. - 11th Int. Conf. Intell. Comput. Technol. Autom. ICICTA 2018* **2018**, 101–105, doi:10.1109/ICICTA.2018.00031.

61.   Martínez-Rodrigo, A.; García-Martínez, B.; Huerta, Á.; Alcaraz, R. Detection of Negative Stress through Spectral Features of Electroencephalographic Recordings and a Convolutional Neural Network. *Sensors* **2021**, *21*, 3050, doi:10.3390/s21093050.

62.   Asghar, M.A.; Khan, M.J.; Fawad; Amin, Y.; Rizwan, M.; Rahman, M.; Badnava, S.; Mirjavadi, S.S. EEG-based multi-modal emotion recognition using bag of deep features: An optimal feature selection approach. *Sensors (Switzerland)* **2019**, *19*, 1–16, doi:10.3390/s19235218.