



Research Paper

An Efficient Multiobjective Feature Optimization Approach for Improving Motor Imagery-based Brain-computer Interface Performance



Sanaz Rezvani^{1,3} , Ali Chaibakhsh^{2,3*}

1. Department of Mechanical Engineering, Faculty of Mechanical Engineering, University of Guilan, Rasht, Iran.

2. Department of Dynamics, Control and Vibration, Faculty of Mechanical Engineering, University of Guilan, Rasht, Iran.

3. Intelligent Systems and Advanced Control Lab, University of Guilan, Rasht, Iran.



Citation Rezvani S, Chaibakhsh A. An Efficient Multiobjective Feature Optimization Approach for Improving Motor Imagery-based Brain-computer Interface Performance. *Caspian J Neurol Sci*. 2024; 10(1):77-86. <https://doi.org/10.32598/CJNS.10.36.461.1>

Running Title Multiobjective Feature Optimization in BCIs

doi <https://doi.org/10.32598/CJNS.10.36.461.1>



© 2018 The Authors. This is an open access article under the [CC-BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license.

ABSTRACT

Background: Applying efficient feature extraction and selection methods is essential in improving the performance of machine learning algorithms employed in brain-computer interface (BCI) systems.

Objectives: The current study aims to enhance the performance of a motor imagery-based BCI by improving the feature extraction and selection stages of the machine-learning algorithm applied to classify the different imagined movements.

Materials & Methods: In this study, a multi-rate system for spectral decomposition of the signal is designed, and then the spatial and temporal features are extracted from each sub-band. To maximize the classification accuracy while simplifying the model and using the smallest set of features, the feature selection stage is treated as a multiobjective optimization problem, and the Pareto optimal solutions of these two conflicting objectives are obtained. For the feature selection stage, non-dominated sorting genetic algorithm II (NSGA-II), an evolutionary-based algorithm, is used wrapper-based, and its effect on the BCI performance is explored. The proposed method is implemented on a public dataset known as BCI competition III dataset IVa.

Results: Extracting the spatial and temporal features from different sub-bands and selecting the features with an evolutionary optimization approach in this study led to an improved classification accuracy of 92.19% which has a higher value compared to the state of the art.

Conclusion: The results show that the proposed improved classification accuracy could achieve a high-performance subject-specific BCI system.

Keywords: Brain-computer interface, Motor imagery, Feature extraction, Feature selection, Optimization

Article info:

Received: 15 Jul 2023

First Revision: 17 Sep 2023

Accepted: 21 Nov 2023

Published: 01 Jan 2024

* Corresponding Author:

Ali Chaibakhsh, Associate Professor.

Address: Department of Dynamics, Control and Vibration, Faculty of Mechanical Engineering, University of Guilan, Rasht, Iran.

Tel: +98 (13) 33693505, Fax: +98 (13) 33693505

E-mail: chaibakhsh@guilan.ac.ir

Highlights

- A novel approach is proposed for improving the feature extraction and selection stages of the machine-learning algorithm applied in a motor imagery-based brain-computer interface (BCI).
- An improved classification accuracy is obtained in this study, which highlights the efficiency of the proposed approach and achieving a high-performance subject-specific BCI system.

Introduction

In the last decade, brain-computer interfaces (BCIs) have emerged as powerful communication systems and enabled humans to interact with their surroundings through a new non-muscular channel using control signals generated from the brain [1]. Amongst all modalities used for acquiring brain activities, electroencephalography (EEG) is the most frequently employed technique [2, 3]. BCIs have enabled users to command external devices merely by imagining the movement in their limbs. The process begins via motor imagery (MI) signals in which sensorimotor rhythms originating from the primary motor cortex provide data are translated to commands and sent to external devices [4].

Designing a motor imagery BCI can be treated as a supervised machine-learning algorithm that aims to identify the user intention by discriminating the classes existing in the data. Applying efficient feature extraction and selection methods is fundamental in improving machine learning algorithms' performance. In classification problems, minimizing the error rate while simplifying the model using a smaller set of features are the key factors to achieving a higher performance system. Various techniques for extracting features have been introduced in the literature. Features are either extracted from the time and frequency domains via spatial techniques or a combination of these domains [5]. Selecting the most informative and discriminative subset of features and then feeding the classifier with these salient features helps the motor imagery BCI to identify the intended motor movements more accurately [4]. Two approaches are primarily adopted in feature selection applications: Filter methods, in which an autonomous assessment criterion is used to evaluate the goodness of a subset of features generated by a search strategy, and wrapper methods, which use the prediction output of a classifier as an objective function to evaluate the feature subset [6]. Since wrapper techniques are developed based on the interaction between the classifier and the features, they generally outperform filter methods in terms of classification accuracy [7, 8].

In motor imagery BCIs, sensorimotor rhythms, which are oscillations observed in mu (8-13 Hz) and beta (13-30 Hz) frequency ranges, contain the most helpful information regarding motor imagery tasks [9]. While designing a machine learning pipeline for a BCI system, selecting the most salient features can be treated with a multiobjective optimization approach in the sense that we are facing two contradictory objectives: Maximizing the classification accuracy (or minimizing the classification error rate) while minimizing the number of features [10]. Obtaining the Pareto optimal solutions that are the trade-off responses of these two conflicting objectives can provide us with a set of non-dominated solutions.

Two main criteria need to be taken care of while choosing a feature selection algorithm: Search strategy and subset quality. The search strategies mainly used in BCI studies are either "sequential," in which features are added or removed successively one at a time, or "heuristic," in which evolutionary algorithms can be stated as the example [11, 12]. Heuristic search algorithms have proved to possess simplicity, flexibility, and high efficiency. To employ an evaluation technique for assessing the goodness of the feature subset, a fitness function can be defined using nature-inspired evolutionary algorithms, which tend to solve real-world problems with complicated nonlinear search spaces more robustly [13, 14].

In this study, the performance of an MI-based BCI is enhanced by improving the feature extraction and selection stages of the proposed machine-learning algorithm [15]. To this end, a multi-rate system for spectral decomposition of the signal is designed, and then the spatial and temporal features are extracted from each sub-band. To maximize the classification accuracy while simplifying the model and using the smallest set of features, the feature selection stage is treated as a multiobjective optimization problem, and the Pareto optimal solutions of these two conflicting objectives are obtained. To explore the feature space and select the salient features, non-dominated sorting genetic algorithm II (NSGA-II), an evolutionary-based algorithm, is used wrapper-based, and its impact on the BCI performance is explored. The

proposed method is implemented on the public dataset: BCI competition III dataset IVa [16]. This study's proposed approach has helped achieve a high-performance subject-specific system tailored to each subject's need. Finally, we thoroughly discussed and compared our results with other studies that tried to solve the classification problem on the dataset in the "results" and "discussion" sections of this research paper.

This study aims to enhance the performance of an MI-based BCI by improving the feature extraction and feature selection stages of the machine-learning algorithm applied to classify the different imagined movements. The contribution of this study compared to the state of the art is an improved classification accuracy value compared to similar studies on the same dataset.

Materials and Methods

Like any other field, the machine learning approach applied in BCI systems usually consists of the following phases: Pre-processing, feature extraction, feature selection, and classification. The detailed approach used in each stage is explained in the following sections.

Dataset and pre-processing

The dataset used in this study was provided by Fraunhofer FIRST, Intelligent Data Analysis Group [17] in 2004. It is known as dataset Iva of BCI competition III. The EEG signals were recorded from 5 healthy subjects: aa, al, av, aw, and ay. Subjects sat in a comfortable chair with arms resting on armrests. This data set contains only data from the 4 initial sessions without feedback. Visual cues were active for 3.5 s to help subjects perform several motor imagery tasks, of which only the information related to the right hand and foot is publicly available. There were two types of visual stimulation: 1) Where targets were indicated by letters appearing behind a fixation cross (which might nevertheless induce little target-correlated eye movements), and 2) Where a randomly moving object indicated targets (inducing target-uncorrelated eye movements). The recording was made using BrainAmp amplifiers and a 128-channel Ag/AgCl electrode cap. Next, 118 EEG channels were measured at positions of the extended international 10/20-system. The data were acquired from 118 channels, and 280 trials for each subject were recorded. The signal was band-pass filtered between 0.05 and 200 Hz and then down-sampled from 1000 Hz to 100 Hz by the provider of the dataset. Common average reference (CAR) was used to remove the common noise and artifacts in the signal.

The formulation for calculating the CAR filter is shown in Equation 1 [18].

$$1. x_i^{CAR}(t) = x_i(t) - 1/N \sum_{(j=1)}^N x_j(t)$$

In the above Equation, $x_i^{CAR}(t)$ is the filtered signal of channel i while $x_i(t)$ and $x_j(t)$ are the potential of i^{th} and j^{th} channels with respect to the reference, respectively. N is the total number of channels.

Feature extraction

In the motor imagery tasks, the subject is asked to imagine performing a specific task for a fixed duration, which in this study is 3.5 s. These fixed time intervals are called trials in the context of the motor imagery tasks. During the data acquisition, the dataset provider labels each trial with a proper class according to the imagined movement by the subject. In the dataset used in this study, the information about the addresses of the start time point of each trial is given. It is also known that 280 trials for each subject are recorded. Hence, by having the information about the start time point and duration of each trial, all the trials rich in the relevant information for the task at hand are separated from the original signal. Then, features are extracted from each trial.

Selecting subject-specific frequency ranges in motor imagery-based BCIs leads to extracting more informative features and designing a BCI system tailored to each subject's need. Hence, in this study, for the feature extraction stage, a multi-rate system for spectral decomposition of the signal is constructed. Then, spatial and temporal features are extracted from each sub-band. To this end, a b -channel filter bank comprising b analysis filters is designed [19]. As mentioned before, μ and β bands are rich in motor imagery-related information. Still, the sub-bands of μ and β with the most helpful information might slightly differ in different subjects. Thus, the frequency band of 8-30 Hz is band-pass filtered into sub-bands with 4 Hz intervals overlapping by 2 Hz using fifth-order Butterworth filters. After constructing the bank of filters, it is applied to each trial. The features from both spatial and temporal domains are extracted from each sub-band, leading to a deeper insight into the characteristics of the EEG data. The architecture of the proposed method in this study is represented in Figure 1.

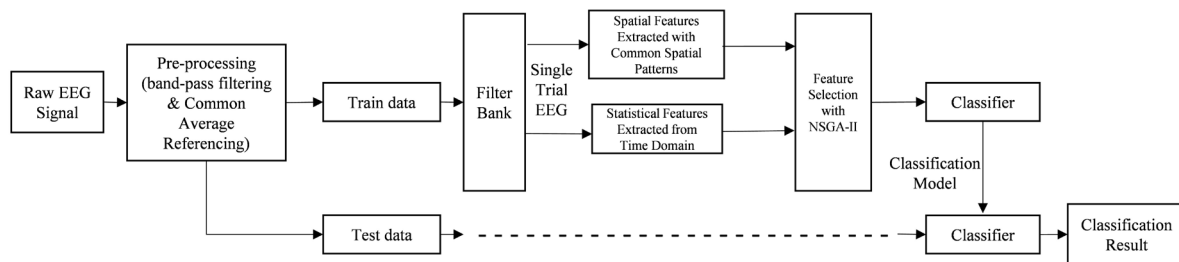


Figure 1. The architecture of the proposed method

Spatial features

The common spatial pattern (CSP) algorithm is one of the most used feature extraction techniques for handling MI signals. By providing the optimized spatial filters, the CSP algorithm helps the multichannel EEG data to be mapped to a new space in which the data variance from one class is maximized. In contrast, the data variance from the other class is minimized [20].

The CSP algorithm can be implemented by extremizing the objective function $J_{CSP}(w)$ as brought in Equation 2 and finding the extremum spatial filters w [10].

$$2. J_{CSP}(w) = \frac{wX_1X_1^T w^T}{wX_2X_2^T w^T} = \frac{w\bar{C}_1 w^T}{w\bar{C}_2 w^T}$$

In the above Equation, X_1 and X_2 are $N \times T$ pre-processed single-trial EEG signals for each class in which N is the number of channels, and T is the number of samples in each trial. X_1^T and X_2^T are the transpose matrices of X_1 and X_2 , respectively. Equation 2 can also be calculated by the use of \bar{C}_1 and \bar{C}_2 which are the average covariance matrices of each trial from classes 1 and 2, respectively. Since $J_{CSP}(w)$ is a Rayleigh quotient, it can be solved via the generalized value decomposition of covariance matrices, which leads to constructing an $N \times N$ projection matrix P_{CSP} . Once the P_{CSP} is obtained, the spatial filters w can be obtained by arranging the eigenvalues in descending order, arranging their corresponding eigenvectors, and then choosing the first m columns and last m columns of the projection matrix [21].

The pre-processed single-trial EEG signals, X_1 and X_2 , are then projected into the low-dimensional spatially-filtered signals Y_1 and Y_2 with the Equation 3:

$$3. Y_i = W^T X_i; i=1 \text{ and } 2$$

, where W^T is the transpose of the $N \times 2m$ matrix containing the first and last m columns of the projection matrix. m can be set to 1 in the same manner as what is used in references [22, 23].

In the final stage of implementing the CSP algorithm, the features are obtained by calculating the variance of the spatially filtered signals by Equation 4 [10]:

$$4. F_i = \text{var}(W^T X_i); i=1 \text{ and } 2$$

Temporal features

After band-pass filtering mu and beta rhythms into different sub-bands, as explained earlier, the statistical measures of the signal are extracted from each trial in the time domain and then, alongside the extracted spatial features, are fed to the next stage of the machine-learning algorithm. Mean, variance, skewness, and kurtosis are the temporal features extracted in this study, giving us detailed information about the time domain features extracted within a specified frequency range.

Evolutionary-based feature selection

In this paper, by using evolutionary algorithms, heuristic search is applied to the problem of feature selection, and a wrapper-based approach is used to select the subset of salient features. To this end, NSGA-II, a multiobjective optimization technique, is used to tackle the feature selection problem. It is an elitist technique with lower computational complexity than similar techniques or its earlier version, NSGA [24]. NSGA-II, a population-based algorithm, starts with a randomly generated population of size n_{pop} and evaluates candidate solutions iteratively to find the global optima in the search space [25].

Similar to the genetic algorithm technique, each member in the population is an n -bit chromosome. In contrast, for the feature selection problem proposed in this study, the algorithm is defined in the binary mode (i.e. chromosome bits only possessing 0 or 1), and n is the total number of features extracted in the previous stage of the machine learning algorithm. Each bit of the chromosome represents the existence or nonexistence of a feature in the feature space by being 1 or 0, respectively.

Since a wrapper approach is used for the feature selection problem in this study, the cost function for evaluating the candidate solutions is defined as the classification error E obtained by feeding the classifier with the selected features where $E=1-A$ and A is the average of 10-fold cross-validation classification accuracy of the train data using the selected features.

In the multiobjective optimization problem stated in this study, the trade-off responses of the two conflicting objectives, maximizing the classification accuracy (or minimizing the classification error rate) while minimizing the number of features, should be obtained. Let the number of selected features be n_f . Once the pairs of candidate solutions with the configuration (n_f, E) are obtained, these solutions should be sorted using the dominance criterion. The “crowding distance” criterion, which is used to sort the population in each generation, and the overview of the domination criterion introduced in [24] used in the NSGA-II algorithm are demonstrated in Figure 2.

Following this stage, the crossover operator on pairs of randomly selected members of the population as parents and the mutation operator on a number of randomly selected chromosomes are applied, and a new population is produced and then evaluated in the same manner. Finally, by merging the initial population, pop , with the population generated by crossover, pop_c , and by mutation, pop_m , and sorting them by applying the dominance criterion once again and then selecting the first n_{pop} members of the population, the Pareto optimal front for the first iteration of running NSGA-II algorithm, as well as

the next generation for use in the next iteration is produced. The flowchart of the feature selection process using the NSGA-II algorithm is shown in Figure 3. In this study, the parameters of the NSGA-II algorithm are set as follows: The population size is 50; the crossover rate and the mutation rate are 0.6 and 0.05, respectively; and the number of iterations is set to 100.

Classification

After the feature selection stage, the best subset of features is fed to a trained classifier that can identify the class of unlabelled trials in the test phase more accurately, leading to the higher performance of the BCI system. This paper uses a classic linear support vector machine as the classifier.

Results

The following actions were conducted to classify the MI signal and determine the subjects’ intention to design a subject-specific brain-computer interface that enables each subject to use the BCI with the best performance possible. The total number of trials for each subject in this dataset is 280. Still, the number of labeled and unlabelled data (train and test data) assigned by the dataset provider [17] for different subjects differs. The percentages of labeled data for subjects aa, al, av, aw, and ay are 60%, 80%, 30%, 20%, and 10%, respectively. The labeled trials and their corresponding spatial and temporal features are used as the training data for the classifier. To tackle the problem of handling the subjects with a high number of unlabelled trials, the employed

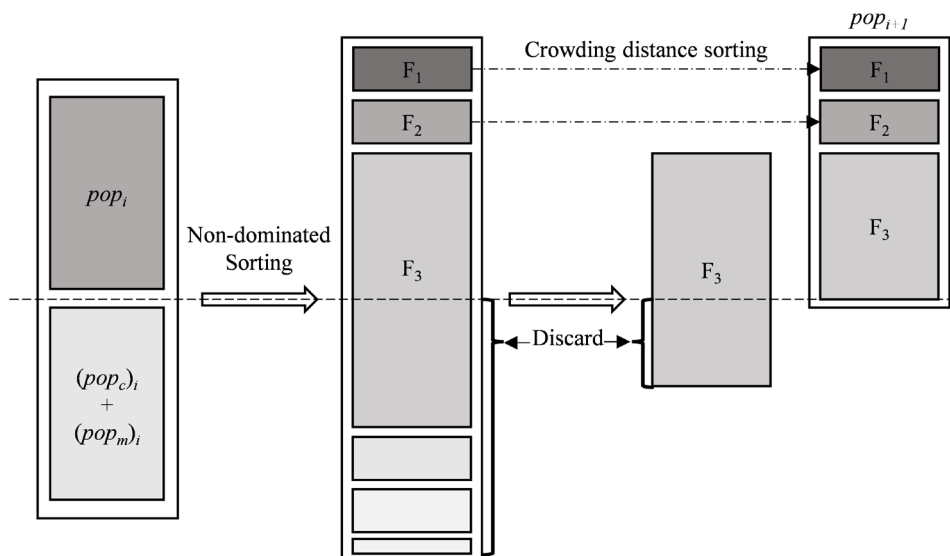


Figure 2. The overview of the domination criterion in the NSGA-II algorithm

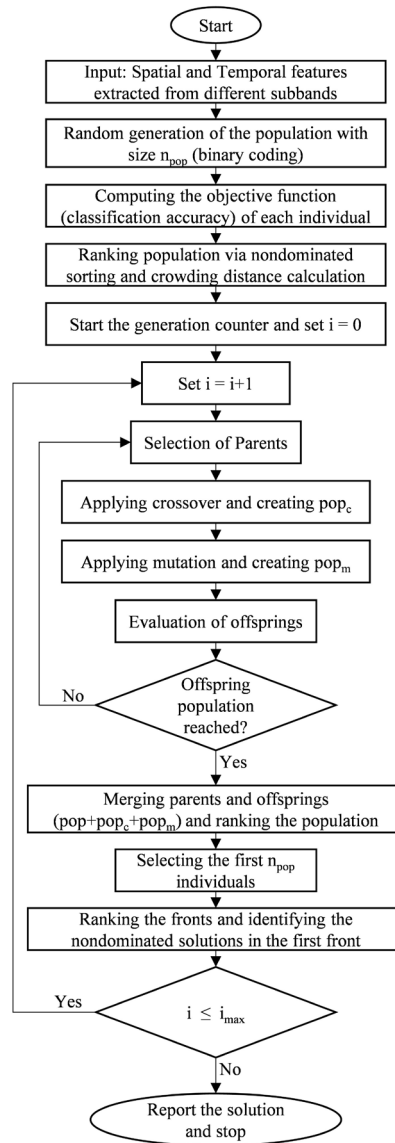


Figure 3. The flowchart of the feature selection process using the NSGA-II algorithm

approach in this study is to construct a pool of features using the correctly classified trials of those subjects with better performance. To this end, in this study, the BCI design process begins with subject al, which possesses the highest number of labeled trials that can be used as the train data. Once the classification of the unlabelled trials of al subjects is done, its data are passed to the next subject to be used as supplementary data. This strategy of data construction ensures that it makes up for the lack of trained data in other subjects. To develop the most efficient configurations to build up the trial pools for different subjects, the results obtained by Dai et al. [26] are considered. However, the final configurations employed in this study differ from those used in the study of Dai et al. [26]. Several studies on this dataset have reported

that after al, the second subject that can achieve a high performance is aw [25–30]. This result is somewhat counterintuitive considering the high proportion of unlabelled trials (80%). A possible reason might be the higher BCI literacy of subject aw compared to other subjects [31]. In this study, for subject aw, the positive impact of constructing a pool of trials can be seen in the 3.57% increase in classification accuracy. That is to say, after feature selection, subject aw achieved the classification accuracy of 94.64% by using subject al as its source of supplementary data. In comparison, this percentage was 91.07% by just using its data. By inspection, the configurations of the supplementary trials for subjects ay, aa, and av are set as al, al+aw, and al+ay, respectively. Sixty-six features are extracted from each trial, of which

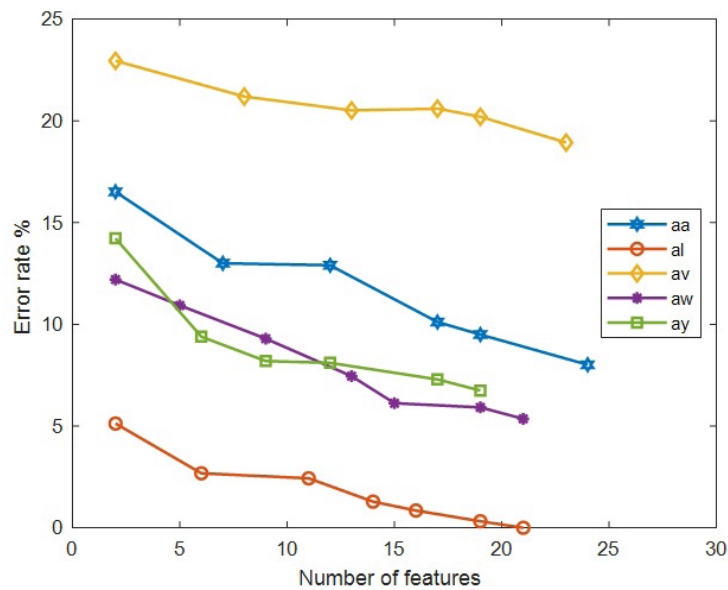


Figure 4. The set of Pareto optimal solutions of the non-dominated sorting genetic algorithm ii for each subject

22 are spatial features extracted by the CSP algorithm, and 44 are the statistical measures of the time domain. Table 1 presents the classification accuracy achieved in different feature extraction and selection stage scenarios.

Figure 4 illustrates the set of Pareto optimal solutions of the NSGA-II algorithm for each subject in this dataset. The final classification accuracies for each subject are selected from the non-dominated solutions in the Pareto front obtained by applying multiobjective optimization for each subject. As seen in Figure 4, the error rate and the number of selected features are negatively correlated, which necessitates making a trade-off between these two conflicting objectives and making a decision to sacrifice the classification accuracy and use fewer features and decrease the computational cost or increase the computational complexity by using more features and achieve higher classification accuracy.

A closer inspection of Figure 4 shows that subject al has achieved the error rate 0 (classification accuracy of 100%) when 21 features are used, which is about less than one-third of the total features. It can also be seen that 21 features for subject aw and 19 features for subject ay provide the minimum error rate for these subjects, and the classification accuracies of 94.64% and 93.25% are achieved, respectively. Subject aa shows a slightly higher error rate than the mentioned subjects, and the classification accuracy of 91.98% is achieved with 24 features. It can be seen that the Pareto front of subject av is situated above all subjects, and this subject possesses the lowest classification accuracy, which is 81.07%. A possible explanation for the low classification accuracy achieved by av might be due to this subject's low signal-to-noise ratio or lack of concentration. Other studies also report this low classification accuracy for subject av on this dataset [25–27, 29, 32, 33].

Table 1. Comparing the results of applying different methods in the feature extraction and feature selection stages

Is Filter Bank used?	Type of Feature	Feature Selection Technique	Number of Extracted Features	Number of Selected Features	Classification Accuracy (%)					Average Accuracy (%)
					aa	al	av	aw	ay	
No	Spatial	-	2	2	58.03	72.23	53.17	69.12	62.81	62.95
No	Spatial+temporal	-	6	6	64.35	85.45	67.31	78.17	72.96	73.65
Yes	Spatial	NSGA-II	22	7	85.97	91.56	78.46	88.63	90.15	86.95
Yes	Spatial+Temporal	NSGA-II	66	21	91.98	100	81.07	94.64	93.25	92.19 (current study)

NSGA-II: Non-dominated sorting genetic algorithm II.

Table 2. Comparing the results of the proposed method in this study with other studies on this dataset

Studies	Technique	Classification Accuracy (%)					Average Accuracy (%)
		aa	al	av	aw	ay	
Ang et al. 2008 [32]	Filter bank CSP	94.00	97.00	86.00	93.00	93.00	88.00
Thomas et al. 2009 [27]	Discriminant filter CSP	90.21	98.86	77.80	97.85	94.23	91.75
Lu et al. 2015 [28]	Structure-constrained matrix factorization	64.21	92.67	60.00	72.58	55.56	68.94
Zhang et al. 2015 [29]	Sparse filter bands CSP	91.64	98.67	77.43	98.03	94.69	92.05
Das et al. 2016 [30]	Subject-specific CFIS	82.14	100	63.27	83.04	60.32	77.75
Dai et al. 2018 [26]	Transfer kernel CSP	68.10	93.88	68.47	90.58	84.65	81.14
Proposed method	Multiobjective spatial-temporal feature optimization	91.98	100	81.07	94.64	93.25	92.19



Table 2 compares the results of the proposed method in this study with other studies on this dataset. Higher classification accuracy achieved with this study's applied approach highlights the proposed method's effectiveness for optimizing the BCI system.

Discussion

One limitation of this study is the inability to provide a comparative analysis of the computational cost of our proposed algorithm with other similar studies on the same dataset. This limitation is attributed to the absence of reported values for this parameter in other comparable studies. Although our research does not conclusively establish the superiority of our algorithm's computational efficiency over existing ones, the heuristic approach suggests the potential for faster performance in our proposed method. Furthermore, the NSGA-II algorithm is an elitist technique with the computational complexity of $O(MO^2)$ (where M is the number of objectives and Q is the population size), which is significantly faster than the similar techniques or its earlier version NSGA with the computational complexity of $O(MO^3)$ and this can lead to a quicker approach compared to the other algorithms used in the similar studies [24].

Another limitation is that the proposed algorithm in this study is applied to the signals of the healthy subjects existing in a public dataset and not to subjects with other brain health conditions. Thus, future research should assess the performance of our proposed algorithm on individuals with brain disorders, namely post-stroke patients.

Conclusion

This study introduces a novel methodology to enhance the feature extraction and feature selection stages within a machine learning algorithm employed in an MI-based BCI. The approach used in this study emphasizes the design of a BCI system that is highly customizable to individual subjects, accounting for subject-specific frequency ranges in motor imagery-based BCIs. Spatial and temporal features were extracted from each sub-band by designing a multi-rate system for spectral decomposition. Afterward, the feature selection stage was treated as a multiobjective optimization problem, prioritizing maximal classification accuracy, model simplification, and using the smallest feature set. Then, the Pareto optimal solutions for these conflicting objectives were successfully determined. Furthermore, exploring feature space and selecting salient features were accomplished using NSGA-II in a wrapper-based manner, with a detailed examination of its impact on BCI performance. Application of the proposed method to the BCI competition III dataset IVa yielded significantly improved classification accuracy compared to previous studies on the same dataset. This outcome highlights the efficiency of our approach, demonstrating its potential for achieving a high-performance, subject-specific BCI system.

Ethical Considerations

Compliance with ethical guidelines

Since the dataset used in this article was already publicly available on the BCI Competition webpage, getting the ethical approval was not required.

Funding

This work was supported by the [Cognitive Sciences and Technologies Council of Iran](#) (Grant No.: 9265).

Authors contributions

Conceptualization and methodology: Sanaz Rezvani and Ali Chaibakhsh; Analysis and validation, writing, reviewing, and editing: Sanaz Rezvani; Supervision and project administration: Ali Chaibakhsh.

Conflict of interest

The authors declared no conflict of interest.

Acknowledgements

The authors would like to thank the [Cognitive Sciences and Technologies Council of Iran](#) for the financial support.

References

- [1] Khan MA, Das R, Iversen HK, Puthusserypady S. Review on motor imagery based BCI systems for upper limb post-stroke neurorehabilitation: From designing to application. *Comput Biol Med.* 2020; 123:103843. [DOI:10.1016/j.compbiomed.2020.103843] [PMID]
- [2] Lazarou I, Nikolopoulos S, Petrantonakis PC, Kompatsiaris I, Tsolaki M. EEG-based brain-computer interfaces for communication and rehabilitation of people with motor impairment: A novel approach of the 21st century. *Front Hum Neurosci.* 2018; ;12:14. [DOI:10.3389/fnhum.2018.00014] [PMID]
- [3] Abiri R, Borhani S, Sellers EW, Jiang Y, Zhao X. A comprehensive review of EEG-based brain-computer interface paradigms. *J Neural Eng.* 2019; 16(1):011001. [DOI:10.1088/1741-2552/aaf12e] [PMID]
- [4] Padfield N, Zabalza J, Zhao H, Masero V, Ren J. EEG-Based brain-computer interfaces using motor-imagery: Techniques and challenges. *Sensors.* 2019; 19(6):1423. [DOI:10.3390/s19061423] [PMID]
- [5] Radman M, Moradi M, Chaibakhsh A, Kordestani M, Saif M. Multi-feature fusion approach for epileptic seizure detection from EEG signals. *IEEE Sens J.* 2021; 21(3):3533-43. [DOI:10.1109/JSEN.2020.3026032]
- [6] Remeseiro B, Bolon-Canedo V. A review of feature selection methods in medical applications. *Comput Biol Med.* 2019; 112:103375. [DOI:10.1016/j.compbiomed.2019.103375] [PMID]
- [7] Baig MZ, Aslam N, Shum HPH. Filtering techniques for channel selection in motor imagery EEG applications: A survey. *Artif Intell Rev.* 2020; 53(2):1207-32. [DOI:10.1007/s10462-019-09694-8]
- [8] Mandal SK, Naskar MNB. Meta heuristic assisted automated channel selection model for motor imagery brain-computer interface. *Multimed Tools Appl.* 2022; 81(12):17111-30. [DOI:10.1007/s11042-022-12327-y]
- [9] Scherer R, Vidaurre C. Motor imagery based brain-computer interfaces. In: Diez P, editor. *Smart wheelchairs and brain-computer interfaces: Mobile assistive technologies.* Massachusetts: Academic Press; 2018. [DOI:10.1016/B978-0-12-812892-3.00008-X]
- [10] Kee CY, Ponnambalam SG, Loo CK. Multi-objective genetic algorithm as channel selection method for P300 and motor imagery data set. *Neurocomputing.* 2015; 161:120-31. [DOI:10.1016/j.neucom.2015.02.057]
- [11] García-Laencina PJ, Rodríguez-Bermudez G, Roca-Dorda J. Exploring dimensionality reduction of EEG features in motor imagery task classification. *Expert Syst Appl.* 2014; 41(11):5285-95. [DOI:10.1016/j.eswa.2014.02.043]
- [12] Tiwari A. A logistic binary Jaya optimization-based channel selection scheme for motor-imagery classification in brain-computer interface. *Expert Syst Appl.* 2023; 223:119921. [DOI:10.1016/j.eswa.2023.119921]
- [13] Kaveh A, Bakhshpoori T. *Metaheuristics: Outlines, MATLAB Codes and examples.* Cham: Springer International Publishing; 2019. [DOI:10.1007/978-3-030-04067-3]
- [14] Yang XS. *Nature-inspired algorithms and applied optimization.* Cham: Springer International Publishing; 2018. [DOI:10.1007/978-3-319-67669-2]
- [15] Kalashami MP, Pedram MM, Sadr H. EEG feature extraction and data augmentation in emotion recognition. *Comput Intell Neurosci.* 2022; 2022:7028517. [DOI:10.1155/2022/7028517] [PMID]
- [16] BCI Competition Data Sets. BCI competition III Dataset [Internet]. 2004 [Updated 22 May 2005]. Available from: [Link]
- [17] Dornhege G, Blankertz B, Curio G, Müller KR. Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. *IEEE Trans Biomed Eng.* 2004; 51(6):993-1002. [DOI:10.1109/TBME.2004.827088] [PMID]
- [18] Yu X, Chum P, Sim KB. Analysis the effect of PCA for feature reduction in non-stationary EEG based motor imagery of BCI system. *Optik (Stuttg).* 2014; 125(3):1498-502. [DOI:10.1016/j.ijleo.2013.09.013]
- [19] Mertins A. *Signal analysis: Wavelets, filter banks, time-frequency transforms and applications.* Chichester: Wiley; 1999. [Link]
- [20] Koles ZJ, Lazar MS, Zhou SZ. Spatial patterns underlying population differences in the background EEG. *Brain Topogr.* 1990; 2(4):275-84. [DOI:10.1007/BF01129656] [PMID]
- [21] Afrakhteh S, Mosavi MR. Applying an efficient evolutionary algorithm for EEG signal feature selection and classification in decision-based systems. In: Mohamed A, editor. *Energy efficiency of medical devices and healthcare applications.* Massachusetts: Academic Press; 2020. [DOI:10.1016/B978-0-12-819045-6.00002-9]
- [22] Ang KK, Chin ZY, Wang C, Guan C, Zhang H. Filter bank common spatial pattern algorithm on BCI competition IV Datasets 2a and 2b. *Front Neurosci.* 2012; 6:39. [PMID]

- [23] Wang Y, Gao S, Gao X. Common spatial pattern method for channel selection in motor imagery based brain-computer interface. *Conf Proc IEEE Eng Med Biol Soc.* 2005; 2005:5392-5. [DOI:10.1109/IEMBS.2005.1615701] [PMID]
- [24] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput.* 2002; 6(2):182-97. [DOI:10.1109/4235.996017]
- [25] Baysal YA, Ketenci S, Altas IH, Kayikcioglu T. Multiobjective symbiotic organism search algorithm for optimal feature selection in brain computer interfaces. *Expert Syst Appl.* 2021; 165:113907. [DOI:10.1016/j.eswa.2020.113907]
- [26] Dai M, Zheng D, Liu S, Zhang P. Transfer kernel common spatial patterns for motor imagery brain-computer interface classification. *Comput Math Methods Med.* 2018; 2018:9871603. [DOI:10.1155/2018/9871603] [PMID]
- [27] Thomas KP, Guan C, Lau CT, Vinod AP, Ang KK. A New discriminative common spatial pattern method for motor imagery brain-computer interfaces. *IEEE Trans Biomed Eng.* 2009; 56(11 Pt 2):2730-3. [DOI:10.1109/TBME.2009.2026181] [PMID]
- [28] Lu N, Li T, Pan J, Ren X, Feng Z, Miao H. Structure constrained semi-nonnegative matrix factorization for EEG-based motor imagery classification. *Comput Biol Med.* 2015; 60:32-9. [DOI:10.1016/j.compbiomed.2015.02.010] [PMID]
- [29] Zhang Y, Zhou G, Jin J, Wang X, Cichocki A. Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface. *J Neurosci Methods.* 2015; 255:85-91. [DOI:10.1016/j.jneumeth.2015.08.004] [PMID]
- [30] Das AK, Suresh S, Sundararajan N. A discriminative subject-specific spatio-spectral filter selection approach for EEG based motor-imagery task classification. *Expert Syst Appl.* 2016; 64:375-84. [DOI:10.1016/j.eswa.2016.08.007]
- [31] Ahn M, Cho H, Ahn S, Jun SC. High theta and low Alpha powers may be indicative of bci-illiteracy in motor imagery. Hu D, editor. *PLoS One.* 2013; 8(11):e80886. [DOI:10.1371/journal.pone.0080886] [PMID]
- [32] Ang KK, Yang Chin Z, Zhang H, Guan C. Filter Bank Common Spatial Pattern (FBCSP) in brain-computer interface. Paper Presented at: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). I01-08 June 2008; Hong Kong, China. [DOI:10.1109/IJCNN.2008.4634130]
- [33] Baig MZ, Aslam N, Shum HPH, Zhang L. Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG. *Expert Syst Appl.* 2017; 90:184-95. [DOI:10.1016/j.eswa.2017.07.033]