

An Implementation Of Machine Learning Algorithms For Feature Optimization And Classification Of Eeg Data

Renuka Suryawanshi¹, Dr. Sandeep B.Vanjale²

¹Ph.D Research Scholar, College of Engineering, BV (DU) COE, Pune, Maharashtra, India.

²Professor, College of Engineering, BV (DU) COE, Pune, Maharashtra, India.

Abstract

One of the most widely used fields in the field of brain conditioning is the processing of electroencephalograms, which is a type of signal that can be used to detect and classify disease. It is very challenging to determine the network of signals that are involved in this process. Early studies on the detection of brain disorder focused on the use of various classification techniques, such as feature extraction and feature selection, to improve the accuracy of the prediction. However, this method has drawbacks in addition to being inefficient, it also requires a large number of features and an increased complexity. Over the past decade, the field of machine control has been heavily studied using brain computer interfaces, which are based on electroencephalography. However, most research on this subject is based on event-based methods. Even though bound methods are useful in medical applications, they can limit the operational capabilities of BCI control systems. This paper presents an initial step in the development of a framework that is designed to improve the accuracy of the classification process by using a combination of features. The framework is designed to improve the accuracy of the classification process by using a combination of features. The first step is to introduce a generational genetic algorithm that is used to select the optimal features for the classification of abstract thought electroencephalograms. Machine learning has been widely used in the field of electroencephalography to perform various tasks, such as analyzing the signals involved in the classification of seizures. One of the most popular algorithms used in this process is the artificial neural network. Unfortunately, its performance might degrade due to certain features. The design of a neural network model should involve the selection of its features. This step is very important because the performance of the network depends on the algorithm that is used to train it. Another technique that is commonly used for training neural networks is backpropagation. Unfortunately, this method can get stuck in a set of biases and weights that are not ideal for its training. Due to the complexity of the training and feature selection process, it is important that the techniques used for training and feature selection are thoroughly studied. This can be done through the use of robust optimization methods.

Keywords: EEG, Brain signal, optimization, Feature selection, Machine learning

1. Introduction

An electroencephalogram, also known as an electrophysiological process, is a type of measurement that involves recording the electrical activity of the brain using a device known as an electroencephalograph. It can be very challenging to deal with due to the noise it produces. Fortunately, medical professionals and scientists have the necessary skills to extract meaningful data from this type of data[1]. Before professionals can view the raw data, they must first process it. The data is presented in a discrete-time time series, which is composed of multiple dimensions. Each dimension is represented by an individual channel in the data.

The sampling rate and total amount of time that was recorded both play a role in determining how many points are included in a time series. These raw signals are utilised very infrequently due to the fact that they may contain traces of electromagnetic noise, DC offsets, and other artefacts. The various steps involved in the process of extracting an electroencephalogram's data are categorized into three main phases: data preprocessing, feature extraction, and recognition. Because of the intricate nature of the data, the final two steps present the greatest amount of difficulty.. After the signal has been enhanced, frequency filters are then introduced to remove the bands of interest[2].

Due to the fact that the evoked response of the human brain occurs in the Theta band, feature extraction is performed so that significant details can be extracted from the collected data. Before the emergence of deep learning, this process was mainly carried out using custom methods. Some of these include linear and spatial filtering techniques. Various general methods are used to extract the data, such as principal component analysis and independent component analysis. On the other hand, more specific methods, such as the energy features and the temporal ones, are commonly utilized. These methods are usually fitted to the specific applications they are used for. For instance, they can identify differences between the different experimental conditions and interpret the data in terms of their predicted behavior.

After the features have been ready, the data collected from the electroencephalogram can be inspected to identify abnormalities, changes in mental status, and average responses. Unfortunately, this process is very time-consuming and costly. It cannot be used in BCI applications as it does not scale accurately. Machine learning, which is a type of artificial intelligence, is an ideal approach to analyzing and improving the data collected from electroencephalography (EEG). It is useful for solving a variety of problems involving the classification of time series. One of the most common approaches to machine learning is known as supervised learning, and it involves using a series of examples as training data and trains a model to identify and predict the patterns in the data[3].

There are many techniques that can be used to classify the data collected from the electroencephalogram. One of the most popular is the classification method, which takes into

account the various classes and regression techniques that can transform the data into a different signal. Some of these include linear techniques, SVM, and neural networks. Due to the complexity of the backpropagation algorithm, it can converge into a set of sub-optimal biases and weights that it cannot escape. This can prevent the neural network from finding a desirable solution. Instead of searching through swarm populations, a neural network should be able to solve specific problems. One of the most popular techniques used in this process is the use of global optimization techniques, such as Grey Wolf Optimizer and Particle Swarm Optimization. These techniques can avoid local optima solutions by ensuring that the initial solutions are random[4].

Through the use of electroencephalography (EEG), researchers can gain a deeper understanding of the human brain and its various aspects, such as emotions and cognition. This technology can also help them identify the underlying causes of human behavior and improve their wellness. In addition to being used for medical diagnosis, it can additionally be utilized to monitor the emotional and physical health of individuals.

Due to the complexity of the data collected by electroencephalography (EEG), it is difficult to interpret. In order to make the analysis more productive overall, a number of techniques have been developed to analyze the data using machine learning. One of the most important issues in classification is the reduction of dimensionality. This can be done by implementing a large number of irrelevant or redundant features in a data set. This can avoid the issue of having more features than patterns in a classification process.

This issue is typically encountered in the development of various applications, such as mass spectroscopy analysis and sequence analysis. This paper reviews the techniques used in feature selection in these applications. It also provides references and analyses for other applications. In order to find sets of genes that are suitable for diagnosis or prognosis, a feature selection procedure is performed in high-dimensional feature spaces.

In order to classify the data collected by electroencephalography, various steps are required. These include identifying the presence of outliers and noise in the features, representing time information in the features, and determining the non-stationarity of the signals. In addition, In most cases, the number of patterns that can be used for training is a significantly smaller number than the number of features that are necessary for the classification process. This is due to the fact that the process of registering the data for each of the different events takes a lot of time. In order to properly classify the signals collected by BCI applications, the number of feature vectors in the data set has to be limited to a few. This issue can lead to the development of a curse-of-dimensionality problem, as the number of patterns that are required to be defined in a classification process increases exponentially. There are two main approaches to reducing dimensionality: the feature space transformation and the selection of features. The former involves either a linear or non-linear transform. This paper aims to provide a framework for the selection of features that will help

improve the performance of a classification process. It also aims to remove redundant or irrelevant features that would make the learning process more difficult.

The concept of the feature selection problem is that it involves searching for the optimal set of features that can be used to solve a clustering problem. The cost function of the selected features is then evaluated to see if they can improve the performance of the classification process. This paper will introduce a wrapper approach for feature selection. It can be used as an alternative to the filter methods. The utility measures of the wrapper approach are related to the features' mutual information and correlation.

2. Related Work

In order to perform denoising on electroencephalography signals, a novel method called wavelet transform (MOFPA-WT) was proposed by Zaid Abdi Alkareem Alyasseri[5]. This method is performed on a standard EEG dataset. The evaluation of MOFPA-WT is carried out using the Keirn EEG dataset, which contains a variety of mental tasks. The performance of the MOFPA-WT algorithm is judged based on its accuracy, false acceptance rate, and true acceptance rate, among other metrics. The proposed method yields the most accurate results, which are attainable through the completion of mental tasks using a geometric figure rotation as the basis.

The proposed feature search method by Tingxi Wen[6], can look for the features of an electroencephalogram signal's instantaneous frequency in response to a given Hilbert transformation. The results achieved by this method are reasonable. The framework exhibits good extensibility because of its use of various classic and modern classifiers. The proposed method can perform well in the classification of two- and three-classification problems and obtained about 99% and 97% respectively. Cross-validation studies conducted on the framework revealed that it can extract significant features from an electroencephalogram signal.

A new dimension reduction method known as Adham Atyabi[7] uses the variations of particle swarm optimization to reduce the number of features and electrodes. The variation in the parameter adjustment and the addition of a mutation operator are used to evaluate the results. The evaluation is carried out according to the degree of performance loss and the selection of the appropriate electrodes. A machine learning algorithm known as an ELM is used to evaluate the set of features and electrodes. The evaluation of the method is performed on a variety of classifiers, such as Perceptron and Polynomial SVM. The results of the study revealed that the variations in PSO can reduce the number of features and electrodes by up to 99%.

In order to find out more about the human brain's status, a proposed data sampling model by Lan Zhang[8] is designed. It can perform various sampling methods. The proposed SEGPA method for performing segmentation and data sampling of an electroencephalogram graph is designed with the help of a weighted network. It can also be used for analyzing the data in a machine control

context. The segmentation and data sampling techniques are combined with other methods such as the normal distribution approximation and the Poisson distribution approximation. The research proposes an efficient method for identifying brain signals and human thinking using frequent patterns generated by the entropy-based pattern recognition system. The obtained framework can be used in robot control or machinery. In the experimental results, the proposed recognition system was able to achieve a 98% accuracy.

A neural network model proposed by Ali Al Bataineh[4] was designed to classify an individual's eye state based on the electrical signals from their brain. In order to improve the model's accuracy, proposed two swarm algorithms: the Grey Wolf Optimizer and the Particle Swarm Optimization. First trained the FFNN model with PSO and GWO instead of using the same features across the different algorithms. Then tested approach against five different machine learning frameworks. These include Logistic Regression, Decision Tree, K-Nearest Neighbors, Support Vector Machine and Gaussian Naive Bayes. According to the results of the study, the FFNN-GWO trainer with the GWO features performed better than the other algorithms. It was able to achieve an accuracy of 97.67%, followed by the FFNN-PSO trainer with the PSO features at 96.40%.

Hemant Choubey[9] proposes a method that combines a check-in function with a masking technique to classify an electrical signal. The integrated K-Means and its corresponding algorithms can then detect if the signal is normal or abnormal. The first step in the process is to extract the signal's features, such as the Signal to Noise Ratio, Standard Deviation, and variance. After that, the features are analyzed and classified according to the Chebyshev distance and the minimum similarity value. The signal can then be classified as normal or abnormal.

Dragi[10] proposes a pair of parallel evolutionary master worker implementations that can be used for the computation of cost functions for different populations. He also proposed a multi-objective procedure for analyzing subpopulations. Experiments were conducted on different benchmarks to evaluate the performance of parallel processing on different tasks, such as the classification of electrical signals from the brain. The results of the studies show that it can improve the quality of the solution and reduce the running time.

Bhavna Kaliraman[11] present a variety of feature extraction techniques that can be used to classify electrical signals. The most common method that achieves the best performance is the autoregressive model. After the extraction of the signals' features, the classification process is performed through feature selection techniques to improve the accuracy of the model. In addition to the conventional methods, present a variety of machine learning algorithms that can be used to classify electrical signals. These include the K-nearest-neighbour and the linear discriminant analysis. The results of the study revealed that the LDA technique can improve the model's accuracy by 99.80%.

Diego Aquino-Brtez[12] presents a method that can be used to improve Deep Learning models' structure and hyperparameters. It can also propose solutions that are composed of different architectures due to their varying layer combinations. The results of the experiments show that the method can outperform the baseline approaches. The proposed framework was evaluated against the baseline conditions of the Deep Learning model and the deep neural network known as EEGNet. The results of the experiments revealed that the method can improve the classification performance by up to 87%.

3. Methods

Feature Extraction and Feature Selection

Only a few of the variables in the dataset are needed to build a machine learning model, and the rest are either irrelevant or redundant. Doing so could negatively affect the model's performance and accuracy. It is very important that the data collected by machine learning is analyzed and selected to meet the specific needs of the users. This process can be carried out through the use of feature selection. When creating a machine learning model, one of the most important factors that the developer must consider is the selection of the various features that will affect the model's performance. This process is carried out in two phases: the first is called feature engineering, and the second is called feature extraction. The goal of extraction and feature selection processes is the same, but the main difference is that the former focuses on creating new features while the latter aims to select the subset of the data that's collected. With feature selection, the model's input variable is reduced by using only relevant data.

Before implementing a particular technique, it is important to understand the need for the technique and the feature selection. In machine learning, it is very important that the input data is good enough to get the most out of it. Collect a lot of data to train model and improve its performance.

The majority of the data collected in machine learning is irrelevant and noisy. This can affect the training process of the model and prevent it from performing well. This is why it is important that the data is removed from the dataset. In order to perform well, the model should be able to select the best features. This process can be done through the use of various techniques such as feature selection. Following feature selection techniques are used

- **Information Gain** - The information gain is a measure of the reduction in entropy that occurs when a dataset is transformed. It can be used to select features by calculating the data gain of each variable.
- **Information Gain Ratio** - The information gain ratio is a tool used to determine which attributes are most relevant to a customer. It can be used to test these attributes near the root of the tree. For instance, one of the attributes that a customer's telephone number provides is a high information gain.

- **Co-relation** - The goal of the correlation feature selection measure is to identify subsets of features that are highly correlated with one another. In this case, the good features are those that are uncorrelated to each other.
- **Principal Component Analysis (PCA)**- A principal component analysis is a process that involves analyzing a large amount of data by taking into account various features and dimensions. It can help improve the interpretability of the data while preserving the most amount of information. The process of transforming a dataset into a new coordinate system is known as PCA, and it can reduce the dimensionality of the data. This technique is useful in identifying clusters of related data points and in plotting the data in two dimensions. The first two components of this process are usually used in conjunction with other methods to analyze the data. A principal component analysis can be used in various fields, such as atmospheric science and microbiome studies.
- **t-distributed stochastic neighbor embedding (t-SNE)** - The technique known as stochastic neighbour embedding makes it possible to visualise high-dimensional data points by assigning a location in a two- or three-dimensional map to each point individually. This allows the data to be viewed more clearly. It is possible to use it to reduce the dimensionality of data in a space with a low number of dimensions. This approach considers all of the different dimensions that an object possesses before modelling those dimensions as a point in either two or three dimensions. Points that are closer together can be used to model similar objects, while points that are further apart can be used to model dissimilar objects.
- **Isomap** - One of the most common low-dimensional embedding methods is the so-called "isomap." This method is used to create a quasi-isometric representation of a set of data points in a complex data structure. It can be used to estimate the intrinsic geometry of the data manifold. The algorithm is very efficient and can be used to solve various data sources and dimensionalities.

4. Machine Learning Classifiers

- **Random Forest (RF)** - A random forest or decision forest is a learning method that can be used in various tasks, such as classification and regression. It can be constructed by training a large number of decision trees at the same time. For instance, in a classification task, the output of the forest is the class that most of the trees are interested in. On the other hand, in a regression task, the average or mean predictions of the individual trees are returned. Decision trees tend to overfitting when training, which is why random decision forests are better than their gradient-boosted counterparts. But, their accuracy is lower than that of other trees due to data attributes.
- **Decision Tree (DT)** - A decision tree is a structure that is made up of internal nodes that each represent a test that is performed on an attribute. Each branch of the decision tree

represents the results of the test, and the leaf nodes represent the class labels that were determined after performing the computation on all of the attributes. The classification rules are represented by the lines that run from the root to the leaf. A decision tree is a tool used in analytical and visual decision analysis to evaluate the expected utility or value of competing alternatives. The influence diagram is also used to visualize the various factors that influence a decision.

- Multi-Level Perceptron (MLP)**- The term "MLP" is used to refer to a fully connected class of artificial neural networks that are feedforward. It can also be used to refer to networks composed of several layers of perceptron. The multiple layers of an MLP are known as nodes. Each node is a neuron that uses a non-linear activation function. This structure can distinguish between linear and non-linear data. It can also be trained using a supervised learning technique known as backpropagation.

5. Results

A number of experiments have been carried out on a variety of characteristics in order to achieve a higher level of precision in its classification. According to the findings of these studies, the method can deliver a higher quality subset of features when appropriate feature selection is performed as compared to when feature selection is not performed.

i. Classification with feature selection

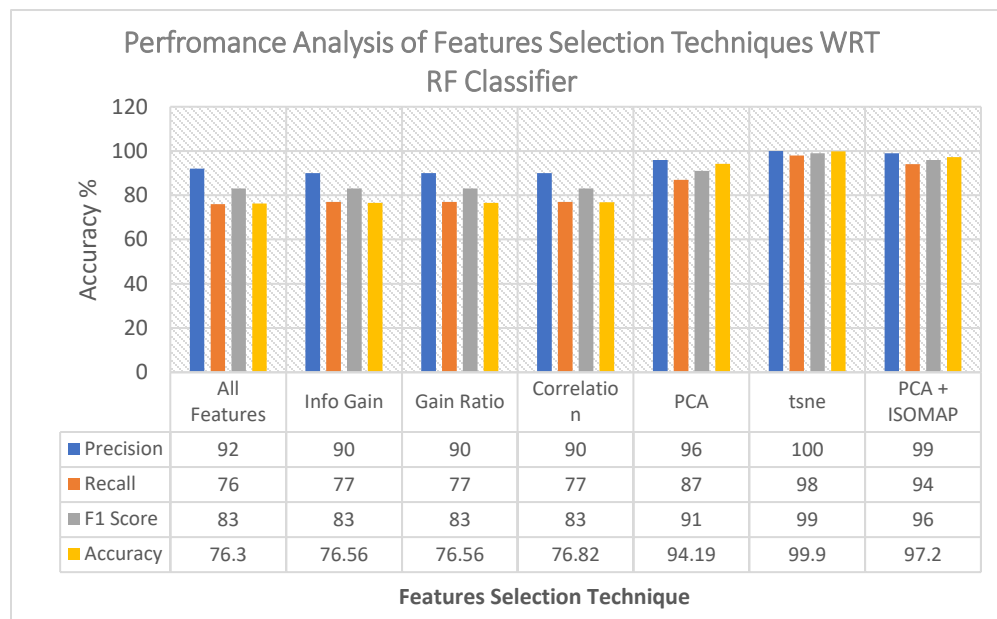


Fig. 1 Result - Random Forest

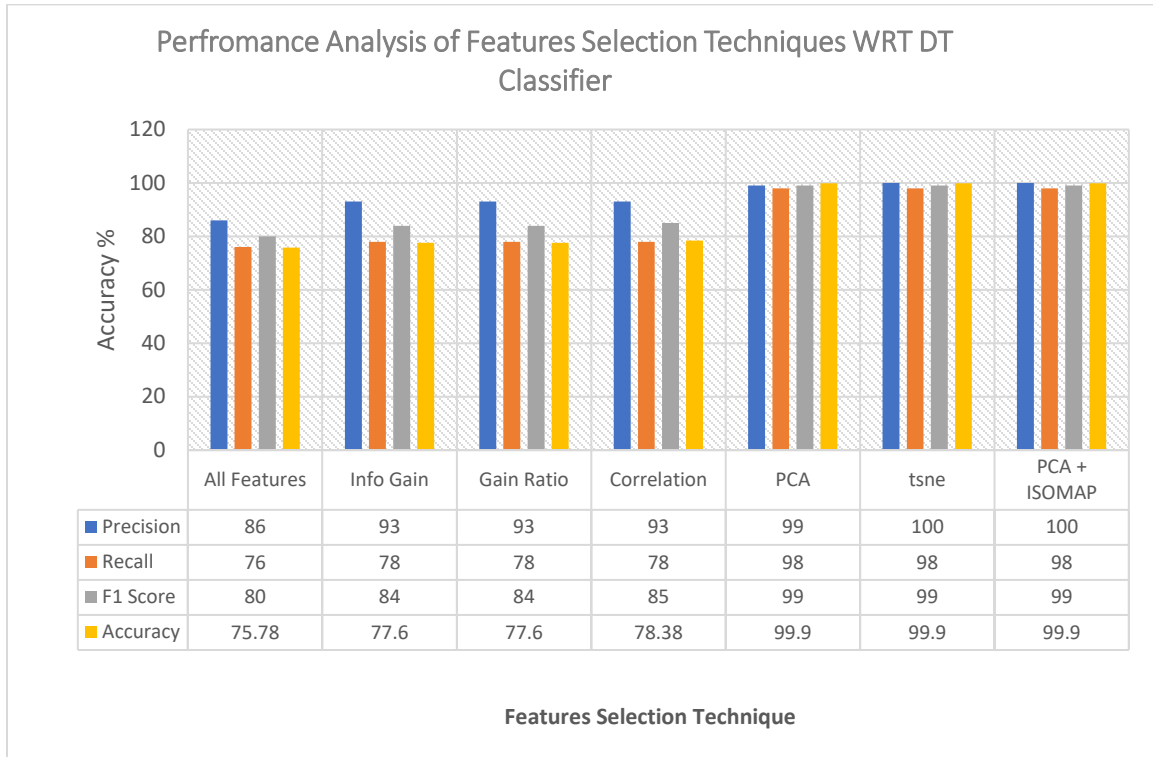


Fig. 2 Result - Decision Tree

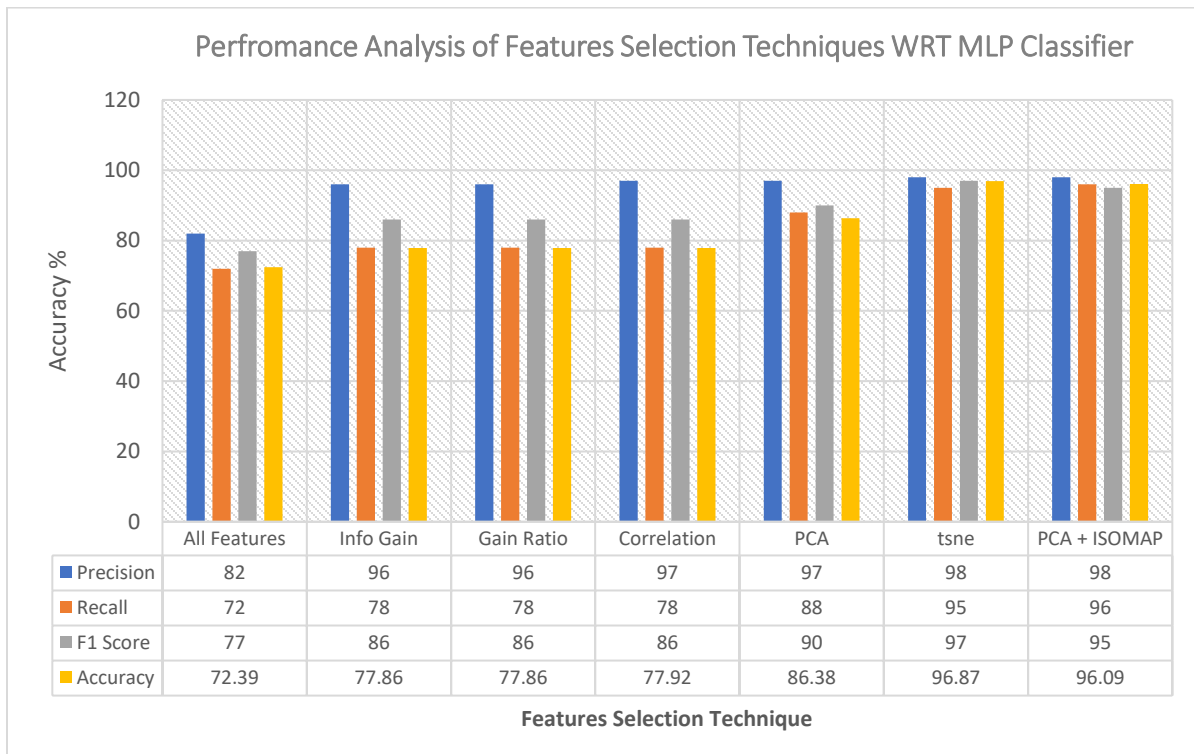


Fig. 3 Result - MLP

ii. Classification without feature selection



Fig. 4 Result - without feature selection

6. Conclusion

This study aims to analyze the various features selection techniques used to eliminate the high dimensionality issue of electroencephalography signals. The complexity of the process makes it more difficult to perform accurate classification. Due to the complexity of classification, it is necessary to improve the performance of the various features. For instance, the PCA + Isomap algorithm performs better than other models.

References

[1] M. Hamada, B. B. Zaidan, and A. A. Zaidan, “A Systematic Review for Human EEG Brain Signals Based Emotion Classification, Feature Extraction, Brain Condition, Group Comparison,” *J. Med. Syst.*, vol. 42, no. 9, 2018, doi: 10.1007/s10916-018-1020-8.

[2] M. M. Eid, F. Alassery, A. Ibrahim, and M. Saber, “Metaheuristic Optimization Algorithm for Signals Classification of Electroencephalography Channels,” *Comput. Mater. Contin.*, vol. 71, no. 2, pp. 4627–4641, 2022, doi: 10.32604/cmc.2022.024043.

[3] S. Chitti, J. T. Kumar, and V. S. Kumar, “EEG Signal Feature Selection Algorithm and Support Vector Machine Model in Patient’s Fatigue Recognition,” *Arab. J. Sci. Eng.*, 2021, doi: 10.1007/s13369-021-06206-1.

[4] A. Al Bataineh and A. Jarrah, “High Performance Implementation of Neural Networks
 3923 <http://www.webology.org>

- Learning Using Swarm Optimization Algorithms for EEG Classification Based on Brain Wave Data,” *Int. J. Appl. Metaheuristic Comput.*, vol. 13, no. 1, pp. 1–17, 2022, doi: 10.4018/ijamc.292500.
- [5] Z. A. Alkareem Alyasseri, A. T. Khader, M. A. Al-Betar, J. P. Papa, O. A. Alomari, and S. N. Makhadme, “An efficient optimization technique of EEG decomposition for user authentication system,” *2nd Int. Conf. BioSignal Anal. Process. Syst. ICBAPS 2018*, pp. 1–6, 2018, doi: 10.1109/ICBAPS.2018.8527404.
- [6] T. Wen and Z. Zhang, “Effective and extensible feature extraction method using genetic algorithm-based frequency-domain feature search for epileptic EEG multiclassification,” *Med. (United States)*, vol. 96, no. 19, pp. 1–17, 2017, doi: 10.1097/MD.0000000000006879.
- [7] A. Atyabi, M. Luerssen, S. Fitzgibbon, and D. M. W. Powers, “Dimension reduction in EEG data using particle swarm optimization,” *2012 IEEE Congr. Evol. Comput. CEC 2012*, pp. 10–15, 2012, doi: 10.1109/CEC.2012.6256487.
- [8] H. L. Zhang, S. Lee, X. Li, and J. He, “Eeg self-adjusting data analysis based on optimized sampling for robot control,” *Electron.*, vol. 9, no. 6, pp. 1–17, 2020, doi: 10.3390/electronics9060925.
- [9] H. Choubey and A. Pandey, “A new feature extraction and classification mechanisms For EEG signal processing,” *Multidimens. Syst. Signal Process.*, vol. 30, no. 4, pp. 1793–1809, 2019, doi: 10.1007/s11045-018-0628-7.
- [10] D. Kimovski, J. Ortega, A. Ortiz, and R. Banos, “Feature selection in high-dimensional EEG data by parallel multi-objective optimization,” *2014 IEEE Int. Conf. Clust. Comput. Clust. 2014*, pp. 314–322, 2014, doi: 10.1109/CLUSTER.2014.6968782.
- [11] B. Kaliraman and M. Duhan, “A new hybrid approach for feature extraction and selection of electroencephalogram signals in case of person recognition,” *J. Reliab. Intell. Environ.*, vol. 7, no. 3, pp. 241–251, 2021, doi: 10.1007/s40860-021-00148-z.
- [12] D. Aquino-Britez et al., “Optimization of deep architectures for eeg signal classification: An automl approach using evolutionary algorithms,” *Sensors*, vol. 21, no. 6, pp. 1–21, 2021, doi: 10.3390/s21062096.