International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.016

Using Statistical Properties and Random Forests, Classification Performance Model Evaluation for the Mental Arithmetic Task-Brain Computer Interface

A.Nirmaladevi
Department of Computer Science and Engineering
SRM Institute of Science and Technology
Katankalathur, Chennai
na7293@srmist.edu.in

R. Vidhya
Department of Computing Technologies
SRM Institute of Science and Technology
Katankulathur, India
vidhyar@srmist.edu.in

Abstract —Ourobjective is to find an efficient method to classify the subject's mental cognitive workload as good or bad by obtaining features that can describe the continuous and underlying temporal dynamics of electroencephalography (EEG) data during the performance of mental tasks. To develop a BCI model that can forecast mental states like good and bad, we explore the ensemble learning technique using classifiers like the random forest classifier. From the alpha, beta, and gamma bands of EEG, the features like mean, root mean square, skewness, mode, data range, interquartile range (IQR), and three Hjorth parameters are extracted to differentiate a signal before-during mental arithmetic task. Our suggested model's analysis and results demonstrate that, when applying these techniques, accuracy is 96%. This model is further utilized for the application of automation in the Internet of Things (IoT).

Keywords—Attention**, **Random Forest Classifier, Support Vector Machine, statistical analysis, Feature Extraction, Feature Selection, Electroencephalography, Brain Computer interface

I. INTRODUCTION

Neuroscientists have recently shown an interest in the creation of Brain-Computer Interface (BCI) [1] devices. The Common technique used for analyzing neural activity is electroencephalography (EEG). Additionally, it might be suggested for the treatment of anomalies, behavioral issues (such as Autism), attention disorders, learning difficulties, language delays, etc. [2]. Depending on the type of activity the EEG wave can be separated into beta, alpha, theta, delta, and gamma waves [3]. These frequency ranges are followed by them.

TABLE 1. THE FREQUENCY RANGES OF EEG WAVES

Waveform	Frequency Range	Activity
Beta	13 - 30Hz	Extremely active interactions and brainactivity
Alpha	8 – 13Hz	extremely calm widening the meditation
Theta	4 – 8Hz	drowsy, falling asleep, and dreaming
Delta	0.1 – 4Hz	A sound slumber without dreams
Gamma	30 – 100Hz	excessive brain activity

A typical BCI paradigm starts with the signal acquisition phase, in which analog brain impulses are gathered and transformed into digital values. Signal capture, data pre-processing, feature extraction, and classification are all processes in any standard BCI model [4]. Noise and artifacts are removed after pre- processing the acquired signals. The process of selecting and extracting certain characteristics from the data for categorization is known as feature extraction. These collected traits are sent into a classifier, which then determines which class they belong to.

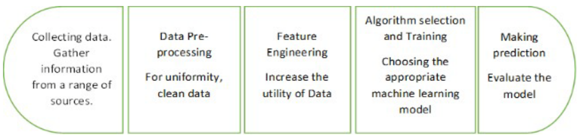

Fig 1. Steps to solve machine learning problems

Classification algorithms can be either supervised or unsupervised, depending on the application. Ensemble learning, as compared to traditionalmachine learning algorithms, is an approach that makes use of several learning algorithms to improve expected outcomes. A group of classifiers is used for learning in ensemble learning, which is a subset of supervised learning. Typically, these ensemble classifiers perform predictions using several straightforward classifiers, such as SVM, naive Bayes, and decision trees and then cast a vote to determine the final class [5]. The classifier used in random forests is made up of several unique classification trees, each of which functions as a separate classifier and is assigned a specific weight in the classification results. By selecting the mode (the output with the most votes) of all the classification outputs from the trees, the overall classification output is calculated [6] The rest of the paper's covers as, in Section II: Related works, in section III: Proposed methodology, which covers details on data collection, features that were extracted, and the techniques that were used to classify. In Section IV, the results and discussion are covered, and the work is wrapped up in Section V.

II. RELATED WORKS

The literature review relates various approaches for EEG signal analysis.From EEG data, Binish Fatimah recovered parameters such as mean, kurtosis, energy entropy, and L2 norms using the rhythms filter [7]. In this

DOI: rp-9788770040723.016

work, the feature extraction and classification methods were performed using SVM, decision trees, quadratic discriminant analysis, and entropy.

Binish Fatimah [8] described a method for identifying mental arithmetic tasks that require solving math problems (serialized subtraction of two numbers). To understand the brain response from a single lead EEG data, the Fourier transform is used. Men and women between the ages of 17 and 26 took part. Participants' ages ranged from 16 to 21 for women. The decomposed signals were filtered for variance characteristics before being classified using the SVM.

Qiang Wang [9] classified real-time EEG waves using multifractal analysis to identify mathematical workloads. In this study, features including power spectrum density and an autoregressive model were employed. The fractal dimension was determined and the attributes were classified using the Support Vector Machine (SVM).

BiswarupGanguly [10] created a categorization of mental arithmetic problems based on EEG to study brain-computer interaction (BCI). EEG data from 36 subjects were recorded, and eight characteristics were extracted from each electrode. These elements were input into the stacked long-short-term memory (LSTM) architecture to create and enhance the brain- computer interface model. FatemaNasrin [11] provides a method for figuring out the functional connectivity between the frontal lobe and pre-frontal lobe areas when young individuals (aged 16 to 20) perform mental arithmetic using single-channel EEG data (subtraction). He concluded that the precision of the Bidirectional Long Short-Term Memory (BLSTM) architecture was considered based on the results of his analysis. With a mean accuracy of 75.88% over 23 channels, this was able to recognize the proper condition of mental arithmetic in 5 seconds. With a significance threshold of less than0.05 in the states of mental calculation and face repose, HodaEdrisAbadi's [12] novel approach involved the extraction of numerous geometric features from Poincare design analysis. This analysis used the crucial comparison t-test to identify variations in brain activity. An artificial neural network has also been used in the two methods to perform autonomous learning and diagnosis (ANN). Electroencephalogram (EEG) data and Bayesian optimized K-Nearest Neighbor were used by LakhanDev Sharma [13] to characterize the mental load and identify the brain's response to stress stimuli (BO- KNN). Entropy-based feature extraction was followed by F-score-based feature selection to increase classification accuracy.

III. Proposed Methodology

3.1 Overview

Data preparation, feature extraction, and classification are the three key components of our suggested methodology. After the data has been gathered, the first stage is data preparation, which is carried out by the EEG device used for data gathering and involves noise reduction and filters. The next stage is feature extraction, where we take the preprocessed data and extract statistical characteristics like mean, root means square, skewness, mode, data range, interquartile range (IQR), and three Hjorth parameters. The final step involves classifying the obtained attributes using a classifier suitable for the collected real-time data. We propose utilizing a random forest classifier in themodel below, which uses an ensemble learning technique, to categorize the supplied data.

3.2 Data Collection

The 36 subjects in the dataset [7] whose signals were captured make up the dataset. 180 seconds before the mental arithmetic task and 60 seconds during the mental arithmetic task were recorded. According to Fig 2 and 3, there are 21 channels in each recording.

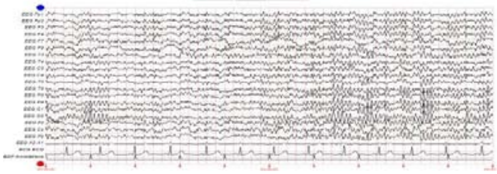

Fig. 2 EEG signals for a 10-second frame of before mental arithmetic calculation

Fig: 3 EEG signals for a 10-second frame duringmental arithmetic calculation

The electrodes are positioned to gather the most signals possible. The job of n-subtraction is given to each participant. Each subject initially receives two variables, x, and y. As many times as feasible, the subject must subtract y from x and recall the outcome. The individual constantly performs this task for one minute. Finally, by computing the remainder, it is determined if the subtraction was performed successfully or not. The sampling frequency was chosen to be 500 Hz

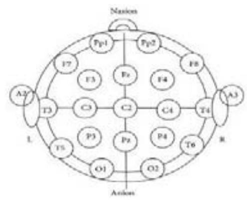

Fig 4. 10/20 International System of electrode placement

3.3 Feature Extraction

A 50 Hz notch filter is used in the pre-processing stage to eliminate the reference DC component. The extraction of characteristics from the alpha, beta, and gamma frequency bands is a step further in the study. Chebyshevbandpass filters with predetermined frequency ranges are used for this. From these filtered data, we have retrieved the features listed below.

1. Mean: The average of all the samples is provided by the signal's mean values.

$$x_{mean} = \frac{1}{N}\sum_{i=1}^{N} x(i) \qquad (1)$$

where N is the signal's overall sample count.

DOI: rp-9788770040723.016

2. Root Mean Square (RMS): The RMS value of XRMS is calculated as follows for a signal x(t)

$$x_{RMS} = \sqrt{\frac{1}{T}\int_{t-T}^{T} x(t)^2} \quad (2)$$

3. Skewness: The asymmetry of the data distribution relative to the central mean is measured. More data is to the left of the mean if the skewness number is negative, and to the right, if it is positive. It is said that the skewness parameter is

$$S = \frac{E(x-\mu)^3}{\sigma^3} \quad (3)$$

Where the sample's mean and standard deviation are$\mu$and σ, respectively (t). 'E' stands for 'expectation'.

4. Mode: The value that appears the most frequently in a dataset is called its mode, or x(t).
5. Data Range: It gives back the difference between the sample dataset's maximum and minimum values x(t).

$$Datarange = xmax - xmin \quad (4)$$

6. Interquartile Range: The interquartile range of the data samples in a time series object is returned by IQR. IQR refers to the difference between an array of random values' upper and lower quartiles (Q3 and Q1, respectively). The quartile measures the median by considering an even dataset of 2n values or an odd dataset of 2n+1 values. The median of the first quartile's values is in Q1, while the median of the third quartile's values is in Q3.

$$IQR = Q_3 - Q_1 \quad (5)$$

7. Kurtosis: Kurtosis is a term used to describe a measure of a random variable's probability distribution.

$$k = \frac{E(x-\mu)^4}{\sigma^4} \quad (6)$$

8. Hjorth Parameters: A time-domain signal's statistical characteristics are shown by Hjorth parameters. Activity, Mobility, and Complexity are the three different parameters [8]. These variables are quite useful when analyzing EEG signals.

a) Hjorth Activity: It shows the variance of any signal x(t)

$$Activity = v(x(t)) \quad (7)$$

b) Hjorth Mobility: Mobility is equal to the square root of the ratio of the variance of the signals and the first derivative's x(t) signals. This variable is proportional to the spectrum's standard deviation.

$$Mobility = \frac{\sqrt{\mathrm{var}\left(\frac{dx(t)}{dt}\right)}}{\mathrm{var}(x(t))} \quad (8)$$

c) Hjroth Complexity: The degree to which the signal shape resembles a pure sinusoid wave is determined by this parameter. If the signal is more like the sine signal, the complexity converges to unity.

$$complexity = \frac{mobility\left(\frac{dx(t)}{dt}\right)}{mobility(x(t))} \quad (9)$$

3.4 Classification

The classification of brainwaves using traditional classifiers has produced positive results in recent years for all forms of BCI data. Some of these include SVM, naive Bayes, k-NN, LDA, and others. Moreover, working with ensemble learning gives us a new way to learn from real-time data. Our method employs binary categorization, allowing us to determine if a person's count quality is good or bad. Our model includes a random forest classifier for classification that employs an ensemble learning strategy for prediction.

3.4.1 Random Forest

The Random Forest Classifier [9] employs a classification technique known as ensemble learning that employs numerous decision trees during the training phase and produces average predictions ofand 90% of the data for the model's training. The two subsections below provide information on the outcomes and performance of our approach.

Many decision tree algorithms are rule-based and depend exclusively on a collection of rules for making predictions about a set of data. In comparison, random forest classifiers find the root node and split the features randomly, as opposed to using the Gini index or information gain to calculate the root node.

1. Precision: It measures the proportion of real positives to all positives.

$$\frac{TP}{TP+FP}$$

2. Recall: The ratio of true positives to all positively classified samples is how it is described.

$$\frac{TP}{TP+FN}$$

3.4.2 Proposed Approach

When compared to decision tree classifiers, random forestclassifiersoperatesimilarly,butwiththe additionofensemblelearning.Thefirstphaseis building many randomly selected decision trees, each of which predicts a certain class based on the information provided. A voting process is used to determine the final class based on the results of the majority after each tree forecasts a class.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

We used 10% of the total data for the model's testing, and 90% of the data for the model's training. The

DOI: rp-9788770040723.016

twosubsectionsbelowprovideinformationontheoutcomesandperformance ofour approach.

4.1. Confusion Matrix

A confusion matrix [16] in the form of a table for a collection of real data and predicted data is used to show a classification model's performance. The classification model's resultant confusion matrix

TABLE 2.CONFUSION MATRIX FOR RANDOM FOREST MODEL IN TEST DATA

Class	Good	Bad
Good	17	2
Bad	2	59

Table: 2 tells that the confusion matrix for random forest in taken for test data, where 17 samples accurately categorized the quality of the count as good, compared to 59 samples that were correctly labeled as bad. 2 samples were incorrectly labeled the count quality as good, compared to 2 samples that were incorrectly classified as bad. Hence, we observe that around 96% of the whole data set is projected into the appropriate classes.

4.2 Performance Evaluation

Precision, recall, accuracy, and other performance metrics are used to evaluate classification performance. Table 3 below is a list of a few of these measurements.

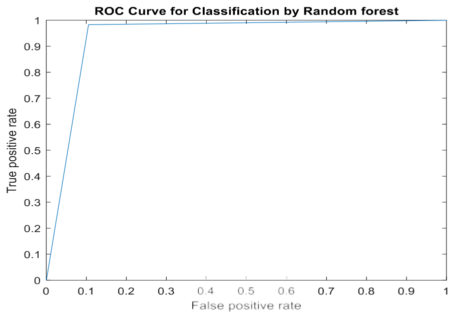

TABLE 3.PERFORMANCE MEASURES

Evaluation Metrics	Values
Correctly Classified Instance	96.20 %
Incorrectly Classified Instances	3.80%
Kappa Statistic	0.8941
Mean absoluteError	0.0380
Accuracy	0.9620
Precision	0.9558
Recall	0.9390
F1- Score	0. 9471
AUC	0.9390

The term "true positives" (TP) refers to samples that were correctly located and anticipated to have high-count quality. The term "true negatives" (TN) refers to samples that were correctly classified as negative and were expected to have poor count quality. False positive samples are those that are mistakenly classified as being good, and false negative samples are those that are classified wrongly as being bad.

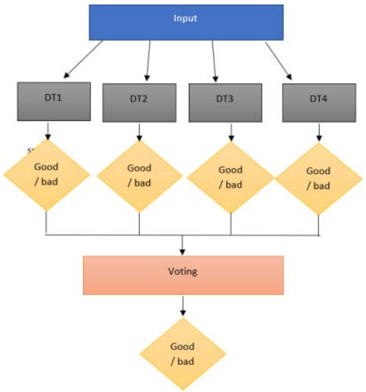

Fig: 5 Structure of Random Forest

The random forest classifier uses an ensemble learning strategy that has a 96% accuracy rate for class prediction.

V. CONCLUSION

In this study, a method for classifying mental states into good and poor categories is suggested. The methodology uses a random forest classifier and statistical feature extraction approach. After attempting to gather EEG data from 36 people, a machine-learning model was created, and the accuracy of this real-time data was 96%. This binary categorization might be used for a variety of things, such in the Internet of Things to turn on and off devices and manage other comparable household equipment by just changing your mindset. The predictive model we developed may be improved and used to control IoT devices more precisely since it gradually learns the users' mental states.

REFERENCES

[1] M. A. P. a. R. M. Kaur, "Technologydevelopment for unblessed people using bci: A survey," International Journal of Computer Applications, p. 40, 2012.

[2] M. I. a. J. R. W. Joseph N. Mak, "Clinical Applications of Brain–Computer Interfaces: Current State and Future Prospects," IEEE Reviews in Biomedical Engineering, vol. 2, pp. 187-199, 2009.

[3] T. U. a. V. Jusas, "Development of a Modular Board for EEG Signal Acquisition," Sensors, p. 18, 2018.

[4] S. Mason and G. Birch, "A General Framework for BrainComputer Interface Design," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 11, no. 1, pp. 70-85, 2003.

[5] C. M. L. A. L. F. A. B. Lotte F, "A review of classification algorithms for EEG-based brain-computer interface.," J Neural Eng, vol. 3, p. 15, 2007 .

[6] L. Breiman, "Random Forests. Machine Learning," Kluwer Academic Publishers., pp. 5-32, 2001.

[7] S. T. I. S. K. K. A. P. M. C. a. O. S. Igor Zyma, "Electroencephalograms during Mental Arithmetic Task Performance," MDPI AG, p. 6, 2019.

[8] Rajesh, M., &Sitharthan, R. (2022). Introduction to the special section on cyber-physical system for autonomous process control in industry 5.0.Computers and Electrical Engineering, 104, 108481.

[9] L. K. K. N. W. H. D. H. FraiwanL,"Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier.," Comput Methods Programs Biomed, vol. 108, no. 1, pp. 10-9, 2012.

[10]Sitharthan, R., Vimal, S., Verma, A., Karthikeyan, M., Dhanabalan, S. S., Prabaharan, N., ...&Eswaran, T. (2023). Smart microgrid with the internet of things for adequate energy management and analysis.Computers and Electrical Engineering, 106, 108556.

[11] M. A. T. C. u. F. D. Method, "Mental Arithmetic Task Classification using Fourier Decomposition Method," IEEE, pp. 0046- 0050, 2020.

DOI: rp-9788770040723.016

[12] O. S. Qiang Wang, "Real-time mental arithmetic task recognition from EEG signals," IEEE Trans Neural SystRehabil Eng., p. 21, 2013.

[13] B. Ganguly, A. Chatterjee, W. Mehdi, S. Sharma and S. Garai, "BiswarupGanguly; ArpanChatterjee; Waqar Mehdi; Soumyadip Sharma; SoumyadeepGarai," IEEE VLSI DEVICE CIRCUIT AND SYSTEM (VLSI DCS), pp. 89-94, 2020.

[14] F. N. a. N. I. Ahmed, "Predicting the Correctness of Mental Arithmetic Task From EEG Using Deep Learning,," IEEE International Conference on Science & Contemporary Technologies (ICSCT), pp. 1- 5, 2021.

[15] M. M. M. M. Abadi HE, "Mental arithmetic task detection using geometric features extraction of EEG signal based on machinelearning," BratislLekListy, p. 6, 2022.

[16] L. &. C. H. &. C. U. &. S. R. &. S. R. Sharma,"Mental arithmetic task load recognition using EEG signal and Bayesian optimized K- nearest neighbor," International Journal of Information Technology, pp. 1-7, 2021.

[17] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," Expert Syst. Appl., vols.1084-1093, p. 32, 2007.