



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Special Issue 2, December 2021

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.569**

# A Statistical Based Approach for Normalization Techniques from EEG Signals in Brain Computer Interface

M.Jeyanthi<sup>1</sup>, Dr.C.Velayutham<sup>2</sup>, Dr. S.John Peter<sup>3</sup>

(Reg No: 19221282282016 ) (Research Scholar), Department of Computer Science, S.T. Hindu College, Nagercoil.

Affiliated to Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India<sup>1</sup>

(Research supervisor), Head and Associate Professor, Department of Computer Science, Aditanar College of arts and science, Tiruchendur, Affiliated to Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India<sup>2</sup>.

(Research Joint Supervisors), Head and Associate Professor, Department of Computer Science, S.T. Hindu College, Nagercoil. Affiliated to Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India<sup>3</sup>

**ABSTRACT:** In Modern Computing technology, BCI research is used for translating the neuronal commands into control signals through classifying the EEG signals patterns of various mental tasks (e.g. imagined left, right hand and foot movements). Data normalization technique is used for improving data quality and subsequently the performance of machine learning algorithms. In this paper, We propose a statistical normalization methods such as Standard Score, Studentized Residual, Standardized Moment, Coefficient of Variation Normalization technique. At present, Statistical measure of feature extraction methods were used to extract the raw EEG data. Afterwards, feature normalization technique were used for reducing EEG variability with rest state of EEG signals. Then, the Chi-Square method were used for feature selection to reduce the number of features for improving the accuracy. Finally, the features are classified using Random Forest. It is observed that the outcome shows that the performance is much more better accuracy than the recently developed algorithms.

**KEYWORDS:** Brain Computer Interface, Feature Extraction, Normalization, Classification

## I. INTRODUCTION

Brain Computer Interfaces (BCIs) is a technology that became a promising conduits for individuals with disabilities and enhance the capability of human brain activities to interact with its external and internal environments. An EEG signals processed onto a device in a readable commands and be analyzed for various patterns. The electroencephalographic (EEG) activity are connected to surroundings without movement of peripheral nerves and muscles to control the real world environment.

The main Contribution of the paper is as follows: In this paper, one of the widely used statistical features is first applied for the feature extraction of BCI Dataset. The proposed method based on four statistical facts such as feature extraction, Normalization, Feature Selection and Classification method are utilized. Based on the first fact, Feature Extraction techniques have been applied to Raw BCI Dataset that is responsible for extracting the useful features. Based on the second fact, Normalization technique is carried on extraction values. Based on the third fact, the extracted features are selected using Feature Selection Methods. Based on the fourth fact, the selected features are classified by Random Forest Classifier with ten fold cross validation. Finally, the results can be obtained.

The organization of the paper is as follows: Section 2 specifies the Related Works of the recent papers, Section 3 specifies the dataset description, Feature Extraction method discussed in Section 4, Normalization techniques are discussed in Section 5 Features selection methods are discussed in Section 6. The Experimental Results are provided in Section 7.

Analysis and Discussion of the results are provided in Section 8. This paper concludes in Section 9.

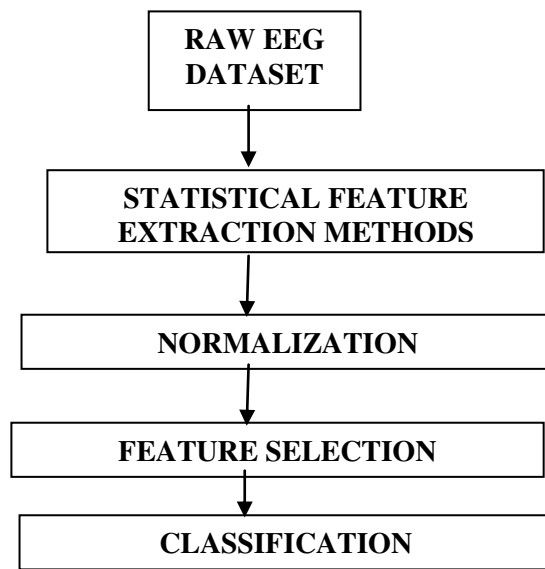


Figure1: Flowchart of the Proposed Work

## II. RELATED WORK

The data sets of EEG data were recorded from several healthy subjects. The cue based BCI paradigm consisted of two or three, namely the imagination of movement of the left hand (LH), right hand (RH) and both feet (F). Several sessions on different days were recorded for some subjects, the data of each session was stored in one data file respectively. In this work we consider only the two class datasets.

The subjects were sitting in a comfortable armchair in front of a monitor. At the beginning of a trial, the screen is blank. After two seconds ( $t=2s$ ), a cue in the form of an arrow pointing either to the left, right or down (corresponding to three classes of LH, RH and F) appeared and stayed on the screen for a specific duration (3-10 sec). This prompted the subjects to perform the desired motor imagery task. The subjects were requested to carry out the motor imagery task until the cue disappeared from the monitor and try to avoid the eye blinking or eye movements during the imagination. A 2 seconds break followed when the cue is disappeared. This procedure is repeated 30-100 times for each run with the random cue sequence. The paradigm is illustrated in Figure 2. For each subject, the first run is called initialization procedure which only presents the cues without any feedback. Based on the online BCI classifier trained on the EEG data recorded from the initialization run, the system is able to present the system feedback online by several red bars representing the classification output for left & right hand and feet commands. Meanwhile, the EEG data with class labels are recorded. The experiments conducted in different days for the same subject are called different sessions.

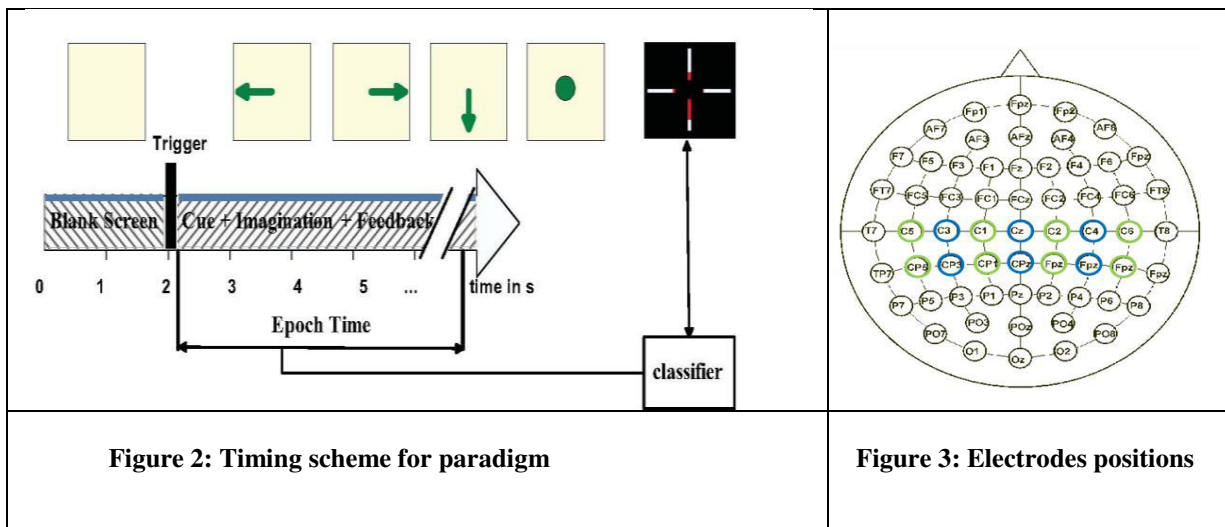


Figure 2: Timing scheme for paradigm

Figure 3: Electrodes positions

### III. METHODOLOGY

#### Data recording

In this data sets, the two devices of g.tec (g.USBamp) and Neuroscan (SynAmps RT) were used for recording the EEG signals. The EEG signals were band-pass filtered between 2Hz and 30Hz with sample rate of 256Hz and a notch filter at 50Hz was enabled for g.tec whereas the band-pass filter between 0.1Hz and 100Hz with sample rate of 250Hz was applied for Neuroscan device. The signals are measured in  $\mu V$  and V for Neuroscan and g.tec respectively. The number of electrodes was different in the data sets, the configuration of 5, 6 and 14 channels were used in different data sets. This aims to develop the BCI system with electrode number as fewer as possible. The electrodes montage is shown in Figure 3. The green and blue electrodes were used in data set with 14 channels, which mainly focus on the motor and sensorimotor area, the blue electrodes were used in data set with 6 channels. For 5 channels data set, the electrodes of C3, Cp3, C4, Cp4, Cz were used to record the EEG signals.

#### Data files and format

The EEG data stored in each data set is organized in segment structure, each segment represents a single trial with one specific class label. The variables in each data set are:

- EEGDATA: The 3-way array with size of [channel x time x trial]. 'Channel' denotes the number of electrodes, 'time' is the duration of each imagination task and 'trial' is the number of motor imagery tasks performed in this session.
- LABELS: vector of target classes (1,2,3) corresponding to each trial in variable 'EEGDATA'. The length of LABELS equals to the length of 3rd-mode of variable 'EEGDATA'.
- Info: structure providing additional information with fields
  - S-rate: sampling rate.
  - class: cell array of the names of the motor imagery tasks.
  - channels: cell array of channel labels.

Table 1 provides the information for each data set file including subject ID, motor imagery class, channel number, duration of each imagination task, trial number, sample rate, device name and the 10x10 folder cross validation performance (accuracy  $\pm$  standard deviation) on this data set. Please note that this performance is roughly obtained by basic preprocess, CSP feature extraction and LDA classifier. The complete same method with same parameters are used to test all data sets without selecting the optimal channels, frequency band and feature number for each subject's data set. Therefore, this



15<sup>th</sup> & 16<sup>th</sup> September 2021

Organized by

Department of Computer Science, Parvathy's Arts and Science College, Dindigul, Tamilnadu, India

results can be used as a general comparison between these data sets and is helpful for understanding which data set is the best and which subject is the best. Each file is recorded from an experiment which is independent from others. This means we can obtain more data sets with 2-class mental tasks for subject B and C by extracting the subset corresponding to the first two classes (left hand & right hand) of the same subject. Hence, the performance of first two classes data sets. In Table 1, the column of *class* denotes the different combination of mental tasks:

- LH/RH: 2 classes of left hand and right hand;
- LH/RF: 2 classes of left hand and right foot;

Dataset	Subject	Class	Channel	Duration (sec)	Trial number	10x10 CV (Acc.±std.)	Sample rate	Device
SubB_6chan_2LR	B	LH/RH	6	4s	162	0.88±0.01	250Hz	Neuroscan
SubE_5chan_2LR	E	LH/RH	5	4s	48	0.86±0.02	256Hz	g.tec
SubF_6chan_2LR	F	LH/RH	6	4s	80	0.71±0.03	256Hz	g.tec
SubG_6chan_2LR	G	LH/RH	6	4s	120	0.81±0.01	256Hz	g.tec
SubC_6chan_2LR_s1	C	LH/RH	6	3s	170	0.86±0.01	256Hz	g.tec
SubC_6chan_2LR_s3	C	LH/RH	6	5s	48	0.92±0.01	256Hz	g.tec
SubC_6chan_2LF_s1	C	LH/F	6	3s	330	0.96±0.002	256Hz	g.tec
SubC_6chan_2LF_s2					180	0.91±0.004		
SubC_6chan_2LF_s3					102	0.86±0.02		

Table 1: The detail information of data set

### 1. FEATURE EXTRACTION

Feature Extraction technique is used as an effective way for both to reducing the computational complexity and to reduce the high dimensional data which can be used to increase the classification accuracies. To reduce the high dimensionality of BCI dataset some of the statistical feature extraction such as mean, mode, median, standard deviations, maximum, minimum, sum, range, variance etc. Feature extraction technique is mainly used for the purpose of reducing the dimension

### 2. STATISTICAL NORMALIZATION

Data Normalization is used to reduces the unwanted variation between the datasets and allow the data on different scales. In this paper, we use different statistical normalization techniques such as

Standard Score	$\frac{X - \mu}{\sigma}$	
Studentized Residual	$\frac{X_i - \hat{\mu}_i}{\hat{\sigma}_i}$	
Standardized Moment	$\frac{\mu_k}{\sigma^k}$	
Coefficient of Variation	$\frac{\mu}{\sigma}$	

### 3. FEATURE SELECTION

In feature selection, we aim to select the features which are highly dependent on the response. The chi-squared statistic[1] is a univariate filter that evaluates feature independently. The value of chi-squared is directly proportional to the relevance of the feature with respect to the class. The chi-squared statistic for a feature is measured as:

15<sup>th</sup> & 16<sup>th</sup> September 2021

Organized by

Department of Computer Science, Parvathy's Arts and Science College, Dindigul, Tamilnadu, India

Chi-Squared Test	$X^2 = \sum_{i=1}^V \sum_{j=1}^B \frac{[A_{ij} R_i * B_j]^2}{R_i * B_j / N}$	
------------------	--	--

where  $V$  is the number of intervals,  $N$  is the total instances,  $B$  is the total number of classes,  $R_i$  is the number of instances in the range,  $B_j$  is the number of instances in class  $j_{th}$ , and  $A_{ij}$  is the number of instances in the range  $i$  and class  $j$ .

#### IV. EXPERIMENTAL RESULTS

In our experiments, we focus on evaluating the classification accuracy for BCI dataset. Ten-fold cross validation was applied to improve the training process. In our experimental study, we use statistical based normalization techniques such as Standard Score, Studentized Residual, Standardized Moment, Coefficient of Variation is used. We first evaluated the classification performance using nine BCI datasets using different normalization methods. In order to improve the BCI Classification performances of the proposed method, we attempted to use the tenfold cross-validation in classification algorithms. First, we use statistical methods to extract the raw EEG signals from the BCI Dataset and again the different normalization techniques are used which helps to normalize the data between 0 and 1. This techniques are used to reduce the computational time and improve the classification accuracy.

#### V. RESULTS AND DISCUSSION

Table 4: Classification Measures									
S. No	Data Set	Standard Score		Studentized Residual		Standardized Moment		Coefficient of Variation	
		AC	ER	AC	ER	AC	ER	AC	ER
1	SubC_6chan_2LF_S1	93	07	88	12	91	09	93	07
2	SubC_6chan_2LF_S2	96	04	85	15	92	08	96	04
3	SubC_6chan_2LF_S3	92	08	90	10	91	09	94	06
4	SubC_6chan_2LR_S1	90	10	76	24	77	23	88	02
5	SubB_6chan_2LR	89	11	75	25	75	25	90	10
6	SubE_5chan_2LR	100	02	85	15	84	16	89	11
7	SubF_6chan_2LR	89	11	87	18	86	14	88	12
8	SubG_6chan_2LR	96	04	88	12	91	09	89	11
9	SubC_6chan_2LR_s3	100	20	79	21	82	08	95	05

Table 4. report the classification results obtained by using classification algorithm of nine BCI datasets. It shows the accuracy, Error rate for three classification algorithm validation such as Hoeffding Tree, Decision Stump, Bayes Net classifier. Our proposed method provides better classification accuracy using classification techniques. As we expected, Hoeffding Tree produces the highest classification accuracy for all BCI compared among other algorithm. Figure 4.shows the results for classification accuracy of BCI data sets.

15<sup>th</sup> & 16<sup>th</sup> September 2021

Organized by

Department of Computer Science, Parvathy's Arts and Science College, Dindigul, Tamilnadu, India

## VI. CONCLUSION

Electroencephalogram (EEG) signals are contaminated with a plenty of artifacts and that they are removed by proposed normalized technique. In this current paper, We have proposed the effect of different score normalization techniques of EEG Signals in Brain Computer Interface accuracy is discussed. Feature Extraction, Normalization, Feature Selection and Classification are the important steps in Signal Processing. In classification step, we used Random Forest classifier and the score of the different classifiers are normalized according to Standard Score, Studentized Residual, Standardized Moment, Coefficient of Variation Normalization technique. The performance of the above methods is appraisement on the BCI dataset. The proposed methods gives much more better result compared to other reported techniques. This technique will not work well if the noise level is high as an EEG signal is highly sensitive to artifacts.

## REFERENCES

1. The author would like to thank to Dr. Cichocki's Lab (Lab. for Advanced Brain Signal Processing), BSI, RIKEN collaboration with Shanghai Jiao Tong University, for providing BCI dataset.
2. Muhammad Zeeshan Baig, Nauman Aslam & Hubert P. H. Shum," Filtering techniques for channel selection in motor imagery EEG applications: a survey",pp. 1207–1232,2020.

## BIOGRAPHY



**M.Jeyanthi** was born in 1988 at Tuticorin, Tamilnadu, India. She received the Master of Computer Applications in 2012 from Anna University, Chennai. Currently, She is an assistant professor with the Department of Computer Science in Aditanar College of Arts and Science, Tiruchendur, Tamil Nadu, India. His research interests include medical image processing, data mining, neural network, Artificial Intelligence, fuzzy logic, and rough set.



**Dr.C.Velayutham** was born in 1965 at Thanjavur, Tamilnadu, India. He received the Master of Science in applied mathematics in 1989, and Post Graduate Diploma in computer application in 1990, from Bharathidasan University, Trichy, India. He received his M.Phil in computer science degree from Manonmaniam Sundaranar University, Trunelveli, India in 2002. He obtained his Ph.d (computer science) degree from Periyar University, Salem, India in 2012. Currently, he is an associate prof. in the Department of Computer Science in Aditanar College of Arts and Science, Tiruchendur. Tamil Nadu, India. His research interests include medical image processing, data mining, neural network, fuzzy logic, and rough set.



**Dr. S.John Peter** is working as Associate Professor in Computer Science at St.Xavier's College, Palayamkottai, Tirunelveli. He received his M.Sc. and M.Phil. degrees in Computer Science from Bharadhasan University, Tiruchirappalli, Tamil Nadu, India. He obtained his Ph.D in MS University. He has published many research papers on clustering algorithms in various Conferences, National and International Journal.



**Impact Factor:  
7.569**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details