IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 787
Research Paper , UGC CARE Listed (Group -I) Journal Volume 11, Iss 11, 2022

Statistical Based Feature Selection Approaches for Motor Imagery EEG Signals in Brain Computer Interface

***M.JEYANTHI**
(Reg No: 19221282282016),
(Research Scholar),
Department of Computer Applications
St.Xavier College (Autonomous) ,
Palayamkottai and Affiliated to
Manonmaniam Sundaranar University
Tirunelveli,Tamil Nadu,India

****Dr.C.VELAYUTHAM**
Research Supervisor
Head and Associate Professor
Department of Computer Science
Aditanar College of Arts and Science
Tiruchendur and Affiliated to
Manonmaniam Sundaranar University
Tirunelveli,Tamil Nadu,India

*****Dr. S.JOHN PETER**
Research Joint Supervisor
Head and Associate Professor
Department of Computer Science
St.Xavier College (Autonomous) ,
Palayamkottai and Affiliated to
Manonmaniam Sundaranar University
Tirunelveli,Tamil Nadu,India

Abstract**—** The fabulous growth of Intelligence Computing leads to the progress of Human Computer Interaction. The Human Nervous System and behavior patterns interactive with intelligence based machine. Now a days, Lot of researchers are concentrate in this area. In this paper, we analysis the existing methodology and also find the feature selection methods. We propose a new Feature Selection method such as Z-Test, Population Variance, Population Standard Deviation, Sample Variance and Sample Standard Deviation. It will outperformed better than some of the benchmark feature selection Methodology.

Index Terms**—**Human Computer Interaction, Feature Selection, Intelligence Computing

I. Introduction

Human Computer Interaction (HCI) has expanded rapidly and steadily for more than four decades. From its origins in human factors engineering and cognitive science into an acclaimed discipline, attracting academics and industry professionals into a multidiscipline dialogue integrating diverse methods, theories and practices. Methodology, theory, and practice in the field of HCI all share similar goals of producing interactive artefacts that can be utilised efficiently, effectively, safely, and with additive user-satisfaction[1].

HCI is simply a hardware and software communications system that enables humans to interact with their surroundings by directly acquiring and analyzing neural signals between the brain and the computer. BCIs are basically devices that translate changes of the neurophysiological activity of the brain into control commands for an application [2]. Unlike the conventional systems which are controlled by computer, the BCI is controlled by human brain signal [3]. The central element of a BCI is the translation algorithm that converts electrophysiological input from the user into output that controls external devices.

The main aim of this paper is to extract features from raw EEG data using statistical feature extraction methods. Then the features are reduced using various feature selection methods. The feature selection methods are evaluated in terms of classification accuracy obtained with SVM Classifiers. Figure1. Shows the organization of proposed BCI system.

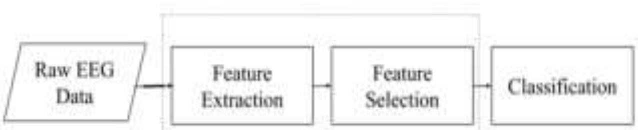

Figure 1:Flowchart for Proposed BCI System

The rest of this paper is organized as follows. Section II focuses on the the literature review, Section III presents the Dataset Description, Section IV describes the Feature Extraction, Section V presents the Feature Selection methods, Section VI describes the Measures of the Performance Evaluation, Section VII describes Results and Discussions, Section VIII presents the Conclusion.

II. LITERATURE REVIEW

Samira Vafay Eslahi et.al.,[4] proposed the GA which performs as a searching engine to find the best combination of the features and classifications. The features used here are Katz, Higuchi, Petrosian, Sevcik, and box-counting dimension (BCD) feature extraction methods. These features are applied to the wavelet subbands and are classified with four classifiers such as adaptive neuro-fuzzy inference system (ANFIS), fuzzy k-nearest neighbors (FKNN), support vector machine (SVM) and linear discriminant analysis (LDA). Due to the huge number of features, the GA optimization is used to find the features with the optimum fitness value (FV). Results reveal that Katz fractal feature estimation method with LDA classification has the best FV.

Izabela Rejer [5] proposed the comparison of two methods of feature selection – PCA (Principal Components Analysis), which is a BSS (Blind Source Separation) method, and method using features from the original feature space. The methods were compared in terms of classification accuracy obtained with neural classifiers using selected features. The comparison was carried out with a data set submitted to the second BCI Competition (data set III – motor imaginary). The data set was recorded from a normal subject (female, 25 years old) which task was to control a feedback bar by means of imagery of left or right hand movements. The data set contains 140 EEG signals, measured over three canals: C3, Cz and C4, sampled with 128Hz and preliminary filtered between 0.5 and 30Hz. In order to perform analysis, the signals were refiltered in 12 different frequency bands and signal power was calculated, separately per each frequency band and each second of the recording.

Michael Schroder et.al.,[6] proposed a feature selection approach based on a genetic algorithm (GA) to pick most promising channels of EEG signals for the classification via support vector machines (SVM). The results of this method are then compared to physiologically motivated feature selection methods and - where applicable - also to the brute force choice of channels.

Aiming Liu et.al.,[7] proposed a novel feature selection method based on the firefly algorithm and learning automata for four-class motor imagery EEG signal processing to avoid being entrapped in the local optimum. After feature extraction using a method of combining CSP and LCD from the EEG data, the proposed feature selection method FA-LA is used to obtain the best subset of features, and the SRDA is utilized for classification. Both the fourth brain–computer interface competition data and real-time data acquired in our physical experiments were used to confirm the validation of the proposed method. Experimental results show that the proposed FA-LA further improves the recognition accuracy of motor imagery EEG, mainly decreases the dimensions of the feature set, and is capable of operating in a real-time BCI system.

Wei-Yen Hsu[8] proposed a BCI system for the classification of MI EEG data. Several potential features such as AAR parameters, spectral power, asymmetry ratio, coherence and PLV are extracted and then combined. Next, the GA is used for feature selection from the feature combination, which significantly enhances classification accuracy. Finally, SVM is used for classification. The experimental results demonstrate that feature selection using GA can further improve performance, and SVM is a satisfactory classifier in the applications of BCI work.

III. DATASET DESCRIPTION

Experimental paradigm

The data sets of EEG data were recorded from several healthy subjects. The cue-based BCI paradigm consisted of two/three motor imagery tasks, 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 computer screen. 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 screen 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 hand, 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.

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 µV 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.

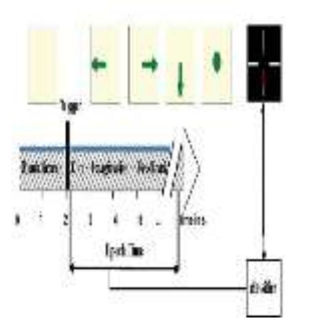

Figure 2: Timing scheme for paradigm

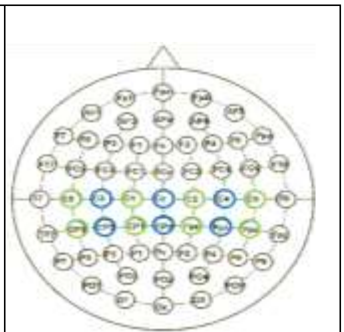

Figure 3: Electrodes positions

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(CN), duration of each imagination task(DUR), trial number(TN), sample rate(SR), device name(D) and the 10x10 folder cross validation performance (accuracy ± 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 results can be used as a general comparison between these data sets and is helpful for understanding which data set is

S. No	Feature Selection Methods	Formula	
1	Z-Test(ZT)	$z = \frac{(x-\mu)}{\sigma}$	(1)
2	Population Variance(PV)	$\sigma^2 = \sum_{i=1}^{n} \frac{(xi-\mu)^2}{n}$	(2)
3	Population Standard Deviation(PSD)	$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{n}}$	(3)
4	Sample Variance(SD)	$s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$	(4)
5	Sample Standard Deviation(SSD)	$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$	(5)

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

Dataset	Subject	Class	CN	DR (sec)	TN	10x10 CV (Acc.±std.)	SR	D
SubC_6chan_2LF_s3	C	LH/F	6	3s	102	0.86±0.02	256 Hz	g.tec

Table 1: The detail information of data set

IV. FEATURE EXTRACTION

Feature extractionis the process of extracting useful information from the raw EEG data.

To compute a feature vector from each of the EEG signals recorded by electrodes located at C3, Cp3, C4, Cp4, Cz, Cpz. Raw EEG dataset which is of size 6 X 102 X 768 where there are 6 channels, 102 trials and data recorded is 256 Hz for 3 seconds. Statistical measures such as variance, mean, standard deviation, maximum, minimum, sum, mode, median, range etcwere used to create features. Which resulting that the size of the dataset become 102 X18.(i,e) 102 objects and 18 features.

V. FEATURE SELECTION

Feature selection is a process that aims to identify a small subset of features from a large number of features collected in the data set [9].

The usefulness of a feature or feature subset is determined by both its relevancy and redundancy. A feature is said to be relevant if it is predictive of the decision feature(s), otherwise it is irrelevant. A feature is considered to be redundant if it is highly correlated with other features. Hence, the search for a good feature subset involves finding those features that are highly correlated with the decision feature(s), but are uncorrelated with each other[10]. In this section, we discuss the FS based on statistical methods such as Z-Test, Population Variance, Population Standard Deviation, Sample Variance and Sample Standard Deviation are used.

Table 2: Statistical Feature Selection Methods

VI. MEASURES OF PERFORMANCE EVALUATION

S. No	Feature Selection Methods	Attributes	Order of Ranking
1	Z-Test	18	f4,f18,f6,f7,f8,f13,f14,f9,f10,f5, f17,f16,f15,f3,f11,f12,f2,f1
2	Population Variance	18	f11,f12,f17,f18,f7,f8,f13,f14,f9,f10,f6,f5,f16,f15,f3,f1,f2,f4
3	Population Standard Deviation	18	f18,f7,f8,f13,f14,f6,f9,f10,f5,f17, f16,f3,f11,f12,f15,f2,f1,f4
4	Sample Variance	18	f6,f18,f7,f8,f13,f14,f5,f9,f10,f2, f15,f17,f16,f3,f11,f12,f1,f4
5	Sample Standard Deviation	18	f9,f10,f 7, f8, f13,f14,f5,f1 , f17,f6,f2,f1,f3,f11,f12,f4,f16,f15

Different measures are used to evaluate the performance of the system. We used tenfold cross validations. Confusion Matrix includes information about actual and predicted classifications applied by a classifier. The data in the matrix is using to evaluate the performance of the classifier. In Table 3 shows the confusion matrix for a two class classifier. It includes TN,TP,FP,FN means True Negative ,True Positive, False Positive and False Negative respectively[11].

		Predicted	
		Negative	Positive
Actual	Negative	TN	FN
	Positive	FP	TP

Table 3: Confusion Matrix

Table 4: Feature Selection Methods with Order of Ranking

VII. RESULTS AND DISCUSSIONS

This section presents the results of experimental studies using data sets extracted from theRaw EEGdata. Only one Dataset are used in our experiment. The data set named as SubC_6chan_2LF_s3and 18 features are used in the corresponding dataset.

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 787
Research Paper , UGC CARE Listed (Group -I) Journal Volume 11, Iss 11, 2022

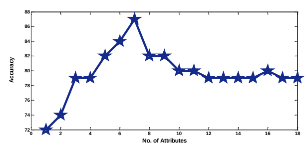

Figure 5: Z-Test

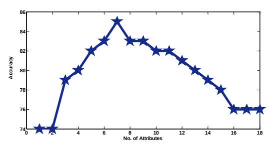

Figure 6: Population Variance

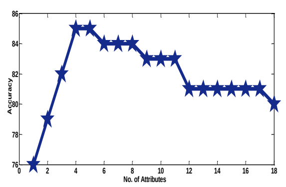

Figure 7: Population Standard Deviation

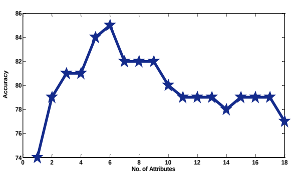

Figure 8:Sample Variance

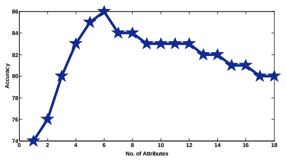

Figure 9: Sample Standard Deviation

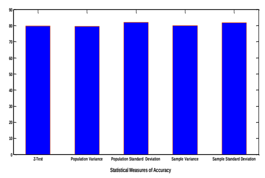

Figure 10: Accuracy of all Statistical Measures

The classification was initially performed on the unreduced data, followed by the reduced data, which were obtained by using the various Statistical Measures of feature selection techniques. Results are presented in terms of classification accuracy. The data presented in Table 5 shows the SVM classification accuracy values of different feature selection methods such as Z-Test, Population Variance, Population Standard Deviation, Sample Variance and Sample Standard Deviation.

The features are arranged according its rank using different feature selection methods, which shows in Table 4. The classification accuracy are evaluated using all the features, eliminating the features one by one from least to high and so on. The recorded classification accuracy values are shown in Table 5.

The feature selection method Ztest, shows the highest classification accuracy 87 with seven featuresf4,f18,f6,f7,f8,f13 and f14. which are demonstrated in Figure 5.
The feature selection method Population Variance, shows the highest classification accuracy 85 with seven features f11,f12,f17,f18,f7,f8 and f13. which are demonstrated in Figure 6.
The feature selection method Population Standard Deviation, shows the highest classification accuracy 85 with have five features f18,f7,f8,f13 and f14. which are demonstrated in Figure 7.
The feature selection method Population Standard Deviation, shows the highest classification accuracy 85 with have six features f6,f18,f7,f8,f13 and f14. which are demonstrated in Figure 8.
The feature selection method Population Standard Deviation, shows the highest classification accuracy 86 with have six features f9,f10,f 7, f8, f13 and f14. which are demonstrated in Figure 9.
In Figure10 shows highest classification accuracy of all the feature selection methods.

Index	Attributes	ZT	PV	PSD	SV	SSD
1	18	79	76	80	77	80
2	17	79	76	81	79	80
3	16	80	76	81	79	81
4	15	79	78	81	79	81
5	14	79	79	81	78	82
6	13	79	80	81	79	82
7	12	79	81	81	79	83
8	11	80	82	83	79	83
9	10	80	82	83	80	83
10	9	82	83	83	82	83
11	8	82	83	84	82	84
12	7	87	85	84	82	84
13	6	84	83	84	85	86
14	5	82	82	85	84	85
15	4	79	80	85	81	83
16	3	79	79	82	81	80
17	2	74	74	79	79	76
18	1	72	74	76	74	74

Table 5: Classification Accuracy of Various FeatureSelection Methods

VIII. CONCLUSION

In this paper we mainly focused on statistical measures using feature selection methods in Brain computer Interface using various statistical approaches such as Z-Test, Population Variance, Population Standard Deviation, Sample Variance and Sample Standard Deviation are clearly analyzed, The SVM Classification algorithm is used to find the classification accuracy. The classification performance is evaluated using confusion matrix with positive and negative class values using ten-fold cross validation,. while compared to all the five methods Z-Test performance is better than the other feature selection methods.

Acknowledgement

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 data-set.

REFERENCES

[1] Konstantinos Koumaditis and Tajammal Hussain," Human Computer Interaction Research Through the Lens of a Bibliometric Analysis", © Springer International Publishing AG 2017 M. Kurosu (Ed.): HCI 2017, Part I, LNCS 10271, pp. 23–37, 2017.

[2] Andrea K., Donatella M., Rüdiger R. and Michael T. (2013), "Facing the Challenge: Bringing Brain Computer Interfaces to End-Users", *Artificial Intelligence in Medicine* 59, 55– www.elsevier.com/locate/aiim Guest Editorial.

[3] Wolpaw J R, Birbaumer N, Heetderkd W J, McFarland D J, Peckham P H, Schalk G, Donchin E, Quatrono L A, Robinson C J, Vaughan T M. (2000), "Brain Computer Interface Technology: a Review of the First International Meeting", *IEEE Transactions on Rehabilitation Engineering*, Vol. 8, No 2, Pp: 164−173.

[4] Samira Vafay Eslahi, Nader Jafarnia Dabanloo, Keivan Maghooli," A GA-based feature selection of the EEG signals by classification evaluation: Application in BCI systems", 16 Jan 2019

[5] Izabela REJER, "EEG Feature Selection for BCI Based on Motor Imaginary Task", Foundations of Computing and Decision Sciences 37(4):283-292 · December 2012.

[6] Michael Schroder, Martin Bogdan and Wolfgang Rosenstiel, Thilo Hinterberger and Niels Birbaumer" Automated EEG Feature Selection for Brain Computer Interfaces", Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering Capri Island, Italy March 20-22, 2003.

[7] Aiming Liu , Kun Chen ID , Quan Liu , Qingsong Ai, Yi Xie and Anqi Chen," Feature Selection for Motor Imagery EEG Classification Based on Firefly Algorithm and Learning Automata", Received: 4 September 2017; Accepted: 3 November 2017; Published: 8 November 2017.

[8] Wei-Yen Hsu," Improving Classification Accuracy of Motor Imagery EEG Using Genetic Feature Selection", Clinical EEG and Neuroscience 2014, Vol. 45(3) 163–168 © EEG and Clinical Neuroscience Society (ECNS) 2013.

[9] M.Jeyanthi, Dr.C.Velayutham, "Analysis of Information Gain Ranking Feature Selection Algorithm Using UCI Machine Learning Datasets", Journal of Emerging Technologies and Innovative Research© *2019 JETIR February 2019, Volume 6, Issue 2:ISSN-2349-5162.*

[10] R. Roselin, K. Thangavel, and C. Velayutham," Fuzzy-Rough Feature Selection for Mammogram Classification", JOURNAL OF ELECTRONIC SCIENCE AND TECHNOLOGY, VOL. 9, NO. 2, JUNE 2011.

[11] M.Jeyanthi, Dr.C.Velayutham, "Machine Learning Verdict of EEG Signals in Brain Computer Interface", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 3, Issue 8, pp.429-441, November-December-2018.

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 currently pursuing Ph.D in the Department of Computer Applications in Manonmaniam SundaranarUniversity,Tirunelveli. 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 obtained his M.Phil (computerscience) degree from Manonmaniam Sunda-ranar University, Tirunelveli, India in 2002. He obtained his Ph.d (computerscience) degree from Periyar University, Salem, India in 2012. Currently, he is an associate professor with the Department of Computer Science, Aditanar College of Arts and Science, Tiruchendur, Tamil Nadu, India. His research interests include medical image processing, data mining, neural network, fuzzy logic, and roughset.

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 Bharadhidasan University, Tiruchirappalli. He also earned his Ph.D. degree from Manonmaniam Sundaranar University, Tirunelveli. He has published many research papers on clustering algorithms in various conferences, national and international journal