Finding Non Dominant Electrodes Placed in Electroencephalography (EEG) for Eye State Classification using Rule Mining

Mridu Sahu[1], N.K.Nagwani[1], ShrishVerma[2]
[1]Department of Computer Science and Engineering, NIT Raipur, C.G., India
[2]Departmet of Electronics and Telecommunication, NIT Raipur, C.G., India

***Abstract*—Electroencephalography is a measure of brain activity by wave analysis; it consist number of electrodes. Finding most non-dominant electrode positions in Eye state classification is important task for classification. The proposed work is identifying which electrodes are less responsible for classification. This is a feature selection step required for optimal EEG channel selection. Feature selection is a mechanism for subset selection of input features, in this work input features are EEG Electrodes. Most Non Dominant (MND), gives irrelevant input electrodes in eye state classification and thus it, reduces computation cost. MND set creation completed using different stages. Stages includes, first extreme value removal from electroencephalogram (EEG) corpus for data cleaning purpose. Then next step is attribute selection, this is a preprocessing step because it is completed before classification step. MND set gives electrodes they are less responsible for classification and if any EEG electrode corpus wants to remove feature present in this set, then time and space required to build the classification model is (20%) less than as compare to all electrodes for the same, and accuracy of classification not very much affected. The proposed article uses different attribute evaluation algorithm with Ranker Search Method.**

***Keywords*—Electroencephalography (EEG); Most Non Dominant (MND); Ranker algorithm; classification; EEG**

I. Introduction

The MND feature subset selection is a part of corpus preprocessing, and it is useful for classification model building as a supervised learning .Classification is one of the task performed by data mining tools and applicable in different area of biomedical electrical devices such as EEG, ECG(Electrocardiograms), EMG(Electromyography), EOG(Electrooculography), Actigraph devices etc. These devices are popular devices for recognizing of different types of disease like Sleep Apnea diagnosis[1] using ECG, driving drowsing using EEG[2],EEG and electromyography (EMG) enable communication for people with severe disabilities [20],muscles activity using EOG[3],and military operation using EEG[21] etc. These are the motivational points for proposed work because the article finds those positions electrode they are less responsible for classification then the removal of those electrodes minimize the size of devices. The present work is performed with EEG electrode data having 16 electrodes and 14892 instances [4,5]. This uses the instance based classifier (K*), because based on statistic of data and nature of data spread over the corpus found it is best among other classifier the result of this present in literature [6, 7], [28], [33], [38]. Method selects either one electrode, two electrode or three electrodes based on how much search space the corpus wants to reduce. Its outcome generated from different attribute selection search with attribute evaluation techniques [8], [37]. Here it is 11 different combination of search with evaluation techniques. Then generating rules using Apriori algorithm [9], it gives frequent electrodes which are placed in ranked as a last four sequences, it also depends how many last feature ranked matrix the corpus wants to create. Here it is 11*4 , where 11 are a Row value and 4 is a column value. Ranker Search with different attribute evaluation algorithms shown in Figure [1].Rankers Algorithm is an algorithm useful for ranking of attributes by their individual evaluation [10]. Here three attribute evaluation methods are defined.

1) **Info Gain Attribute evaluation**: Evaluate the worth of an attribute by measuring the gain ratio with respect to the class.

2) **Classifier Attribute Evaluation**: Evaluate the work of n attribute by using a user specified classifier.

3) **OneR Attribute Evaluation**: Evaluate the work of an attribute by using the oneR classifier.

II. Association Rule Mining

Association Rule Mining is used here for obtaining frequent set they are correlated with each other using support and confidence parameters [11- 13].

Support is define as how frequently a specific item set occur in the data base (the percentage of transactions that contain all of the items in the item set, here the set of items are electrodes present in corpus and the transaction is the different method used for evaluation).

Confidence is the probability that items in RHS (Right Hand Side) will occur given that the items in LHS (left hand side) occurs. It Computed as

Confidence (LHS) =>Support (LHS U RHS)/ Support (LHS) Electrode1 => Electrode2 [0.588, 0.88]

If Electrode1 is selected in MND set, then Electrode2 also selected in MND set if it will satisfies minimum support and minimum confidence value. Left hand side electrode as Antecedent and Right hand side electrode [RHS] as consequent frequency.

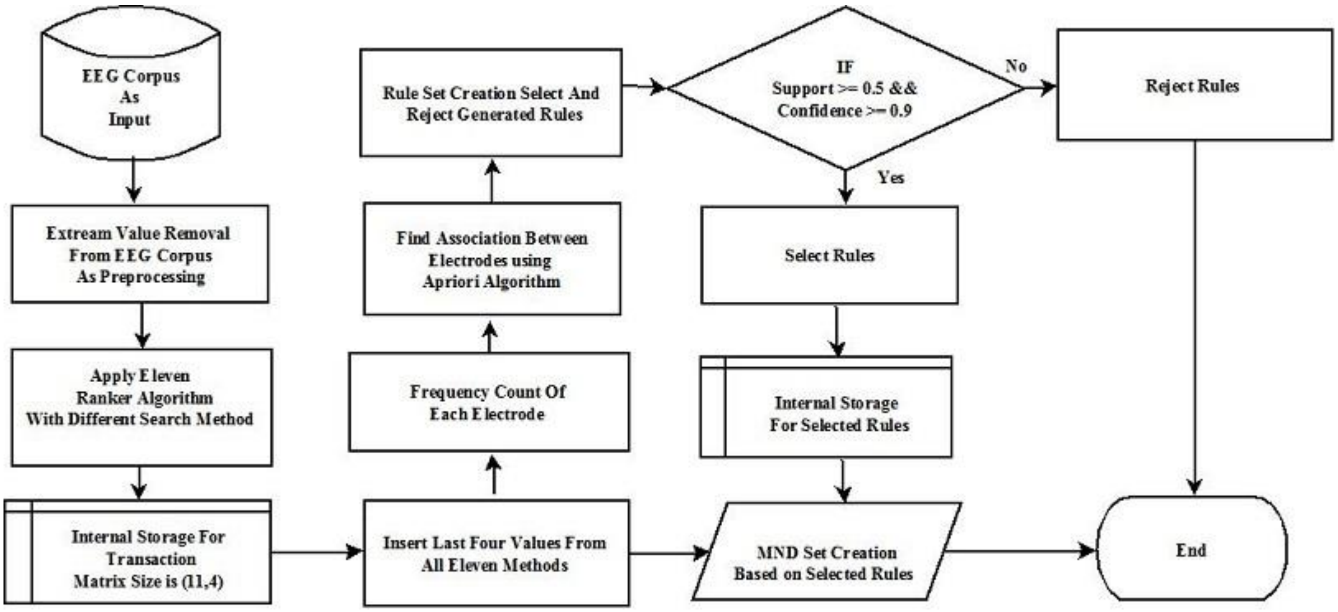

Fig. 1. Flowchart of Proposed Work

III. Electroencephalography (EEG)

EEG is useful for measuring brain activity. During the test very little electricity is passed between the electrodes and skin. EEG usually takes 30-60 minutes. The technician will put a sticky gel adhesive on 16 to 25 electrodes at various spots on our scalp [14]. There are various spatial resolution of EEG systems like 10/20, 10/10, 10/5 systems as relative had surface based positioning system. The international 10/20 system a standard system for electrode positioning with 21 electrodes extended to higher density electrode such as 10/10 and 10/5 systems allowing more than 300 electrode positions [15].

Here the proposed methodology is used in 10/20 system with 16 electrodes (AF3, F7, F3, AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4).

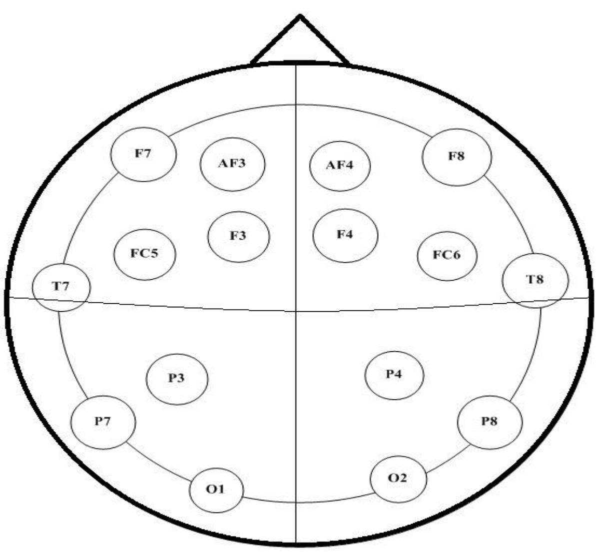

Fig. 2. EEG positioned electrodes 10/20

TABLE I. Search Method used with Different Attribute Evaluators

SEARCH METHOD + ATTRIBUTE EVALUATOR	TRANSACTION
Ranker + InfoGainAttributeEval	T1
Ranker + ChiSquaredAttributeEval	T2
Ranker + ClassifierAttributeEval	T3
Ranker + CorrelationAttributeEval	T4
Ranker + CVAttributeEval	T5
Ranker + FilteredAttributeEval	T6
Ranker + GainRatioAttributeEval	T7
Ranker + OneRAttributeEval	T8
Ranker + ReliefFAttributeEval	T9
Ranker + SignificanceAttributeEval	T10
Ranker + SymmetricalUncertAttributeEval	T11

IV. EEG Corpus

The Corpus consists of 14980 instances with 15 features each (14 features representing the values of Electrodes and one as eye state (Boolean Variable)).Statistical Evaluation finds extreme values present in the corpus, here thirty eight instances **(186, 899, 10387, 10674, 10675, 10676, 10677, 10678, 10679, 10680, 10681, 10682, 10683, 10684, 10685, 10686, 10687, 10688, 10689, 10690, 10691, 10692, 10693, 10694, 10695, 10696, 10697, 10698, 10699, 10700, 10701, 10702, 10704, 10707, 10708, 10709, 11510 and 13180)** declared as extreme values in this, removal of it makes new corpus and it is having 14942 instances . The stored corpus as ordered to able to analyze temporal dependency 8220(55.01%) instances of the corpus corresponds to the eye open and 6722(44.99%) instances to the eye closed. EEG eye state dataset was donated by Rosler and Suendermann from Baden-Wuerttemberg Cooperative State University (DHBW), Stuttgart, Germany [4]. The output of the corpus "1" indicates the eye-closed and "0" indicates the eye-open state.

V. Extreme Value Removal

The extreme value removal is a part of data cleaning step for data mining. The procedure for applying the extreme value theorem is to first establish that the function is continuous on the closed interval [16]. The next step is to determine the critical points in the given interval and evaluate the function at these critical points and at the end points of the interval. If the function f(x) is continuous on closed interval [a, b] then f(x) has both a maximum and a minimum on [a, b] [17]. In proposed method inter-quartile range [IQR] is used for extreme value calculations. IQR is major of variability based on dividing the dataset into quartiles [18].

VI. Feature Subset Selection

Feature Subset Selection is a task of data mining tool[25,26] ,it selects optimal feature subset for classifying the dataset but the literature shows the subset of optimal feature may or may not be optimal[19],[22-24]. The proposed work is searching Most Non Dominant features (MND) from the feature set. This performed by ranker algorithm and with different search methods. The outcome of this step is ranks of electrodes placed in scalp. Proposed work used different 11 algorithms for obtaining the ranks of electrodes (most to least dominant).

TABLE II. Transaction in Matrices with Four Last Dominant Attributes

Transaction	L4	L3	L2	L1
T1	O2	F7	FC5	F3
T2	O2	F7	FC5	F3
T3	FC6	O2	FC5	F7
T4	P7	O1	FC5	T7
T5	F7	AF4	F8	AF3
T6	O2	F7	FC5	F3
T7	F7	FC5	O2	F3
T8	FC6	O2	FC5	F7
T9	F3	F4	O2	P8
T10	P8	O2	F3	F7
T11	F7	FC5	O2	F3

VII. Classification

Classification is the task of data mining and it is a supervised learning. To classify EEG signals, various classification techniques present in literature [34-38]. The instances present in corpus for eye state recognition using EEG, these instances are classified in to two different classes, zero is for eye opened state and one is for eye closed state. The instance based classifier is a type of lazy classifier [27], and proposed method uses K* is a type of instance base classifier, after extreme value removal and attribute selection. The literature shows there are various statistical measures are used for analysis of classification outcomes generated from classification process [29-32].

VIII. Proposed Methodology for MND Set

The proposed methodology is use full for finding non-dominant feature from feature set. If "n" number of features are used for classification of eye state recognition then the space and time requirement is very high but if using less no of features obtained from proposed method then this will save time and space requirement. MND set electrodes are always a most non-dominant electrodes they are less responsible for classification accuracy. The flowchart shows in figure [1], and described steps shows, how to get MND from feature subset results generated from previous step.

S. No.	LHS		RHS	Support	Confidence	Lift
1	**{}**	**=>**	**{F7}**	**0.8182**	**0.8181818**	**1**
2	{FC6}	=>	{FC5}	0.1818	1	1.375
3	{FC6}	=>	{O2}	0.1818	1	1.375
4	{FC6}	=>	{F7}	0.1818	1	1.222222
5	{P8}	=>	{F3}	0.1818	1	1.571429
6	{P8}	=>	{O2}	0.1818	1	1.375
7	**{F3}**	**=>**	**{O2}**	**0.5455**	**0.8571429**	**1.178571**
8	**{F3}**	**=>**	**{F7}**	**0.5455**	**0.8571429**	**1.047619**
9	**{FC5}**	**=>**	**{F7}**	**0.6364**	**0.875**	**1.069444**
10	**{O2}**	**=>**	**{F7}**	**0.6364**	**0.875**	**1.069444**
11	{FC5, FC6}	=>	{O2}	0.1818	1	1.375
12	{FC6, O2}	=>	{FC5}	0.1818	1	1.375
13	{FC5, FC6}	=>	{F7}	0.1818	1	1.222222
14	{F7, FC6}	=>	{FC5}	0.1818	1	1.375
15	{FC6, O2}	=>	{F7}	0.1818	1	1.222222
16	{F7, FC6}	=>	{O2}	0.1818	1	1.375
17	{F3, P8}	=>	{O2}	0.1818	1	1.375
18	{O2, P8}	=>	{F3}	0.1818	1	1.571429
19	{F3, FC5}	=>	{O2}	0.3636	0.8	1.1
20	{F3, FC5}	=>	{F7}	0.4545	1	1.222222
21	{F3, F7}	=>	{FC5}	0.4545	0.8333333	1.145833
22	{F3, O2}	=>	{F7}	0.4545	0.8333333	1.018519
23	{F3, F7}	=>	{O2}	0.4545	0.8333333	1.145833
24	**{FC5, O2}**	**=>**	**{F7}**	**0.5455**	**1**	**1.222222**
25	**{F7, FC5}**	**=>**	**{O2}**	**0.5455**	**0.8571429**	**1.178571**
26	**{F7, O2}**	**=>**	**{FC5}**	**0.5455**	**0.8571429**	**1.178571**
27	{FC5, FC6, O2}	=>	{F7}	0.1818	1	1.222222
28	{F7, FC5, FC6}	=>	{O2}	0.1818	1	1.375
29	{F7, FC6, O2}	=>	{FC5}	0.1818	1	1.375
30	{F3, FC5, O2}	=>	{F7}	0.3636	1	1.222222
31	{F3, F7, FC5}	=>	{O2}	0.3636	0.8	1.1
32	{F3, F7, O2}	=>	{FC5}	0.3636	0.8	1.1

Fig. 3. Rule Generated from Apriori Algorithm

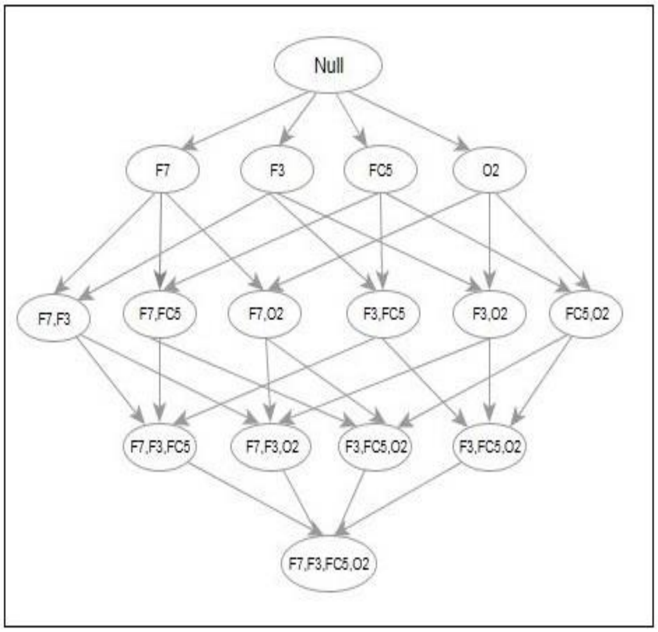

Fig. 4. Lattice occurs by Association Rule Mining

Algorithm 1: Most Non Dominant (MND) Feature Set Generation Algorithm

Input: D, I , T , C, MSupport.

//D = Set of Electrodes.
// I = Total Instances
//T = Transactions
//C = Corpus
// L= Class label {0, 1}
//MSupport=Minimum Support

Output: MND set

Segment(C); // *call this for creating a training, testing and validation set creation.*
For i=1 To 11 do
 For j=1 To 4 do
 T[R] [C] =LRS(C);
// *Call function for last 4 values from different feature ranker search with evaluation techniques*
//Transaction Matrix insertion for Item Set (Electrode) placed last 4 positions.

MND=Apriori(T,Msupport);
//Calling Apriori for frequent set generation for

End

Function Definition for Segment Creation from Corpus

Segment(C)
{
T= 60/100 *(C); *//Training Set Creation*
R= C-T;
Te = 50/100*(R); *//Testing Set Creation*
R= R-Te;
V= R; *//Validation Set Creation*
Return (T, Te, V)
}

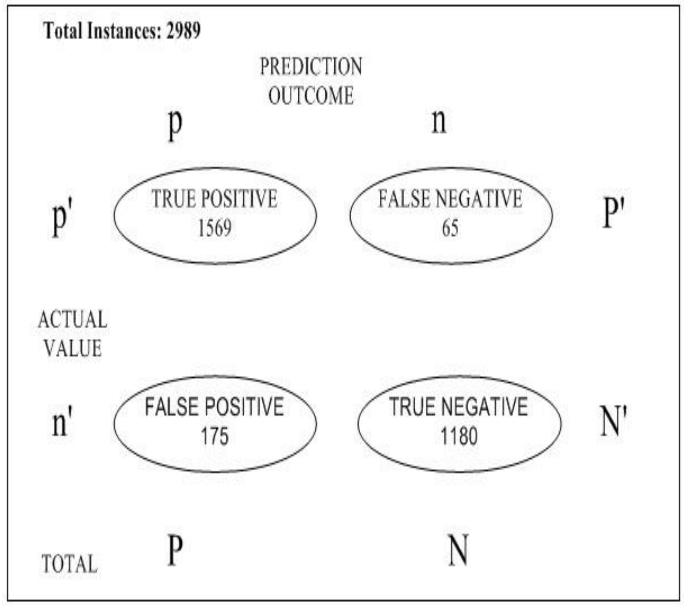

Fig. 5. Confusion Matrix on Removal of F7, FC5, O2

Function Definition for Last Ranked Set (LRS)

LRS(C)
{

For i=1 To 11 do
For j=1 To 4 do

Ran[j] =Ranker (i); *//Last 4 ranked value search and stored in array*
End //for End

Return (Ran[j]);

End
}

Function Apriori Algorithm for Frequent Set Mining

Apriori (T, mSupport)
{
//T is the database and mSupport is the minimum support
F_1= {frequent items};

For (k= 2; F_{k-1}! =∅; k++)
{
Ck= candidates generated from F_{k-1}
//*Cartesian product F_{k-1} x F_{k-1} and eliminating any k-1 size item set that is not frequent*

For each transaction do {
//*increment the count of all candidates in Fk that are contained in T*

F_k = candidates in C_k with minimum Support}

} end for inner for Return $\bigcup_k L_k$; }

IX. Result and Analysis

This study used Ranker Search with Attribute Evaluation technique for MND set creation shown in table[1], then for rule generated using association rule mining this task performed by using Apriori algorithm ,all the generated rules are shown in figure[3],and the lattice shown in figure [4],shows how many frequent set to be considered for rule generation , the rules which is having minimum support and confidence is highlighted in figure[3],this gives frequent items (Electrode) set ,here it is {FC5,O2,F7}. This set declared as MND set, removing of this electrodes from EEG corpus sufficiently decrease the space and time requirement to built the classification model. The accuracy towards the classification changed very less and this analysis outcome shown in table [3], figure[6]. The Confusion matrix shown in figure [5] and ROC curve shown in figure [7], evaluate the classifier performance here the classifier is Instance based classifier (K*), the classification accuracy is computed and it is mapped in table [3].

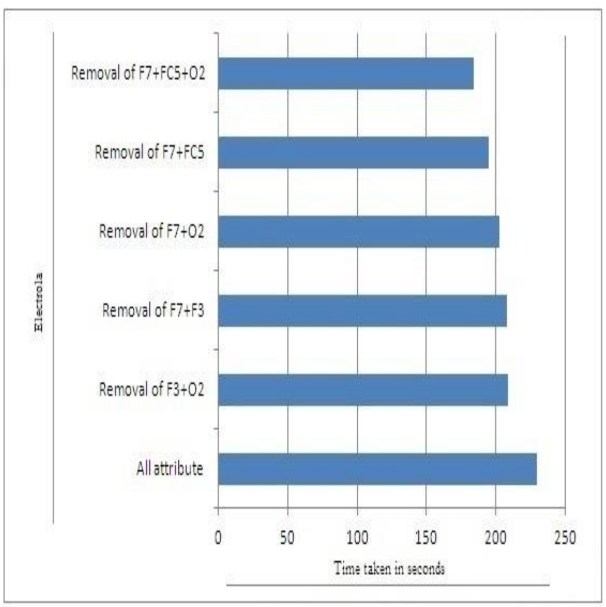

Fig. 6. Time duration with Removal of Different Attributes

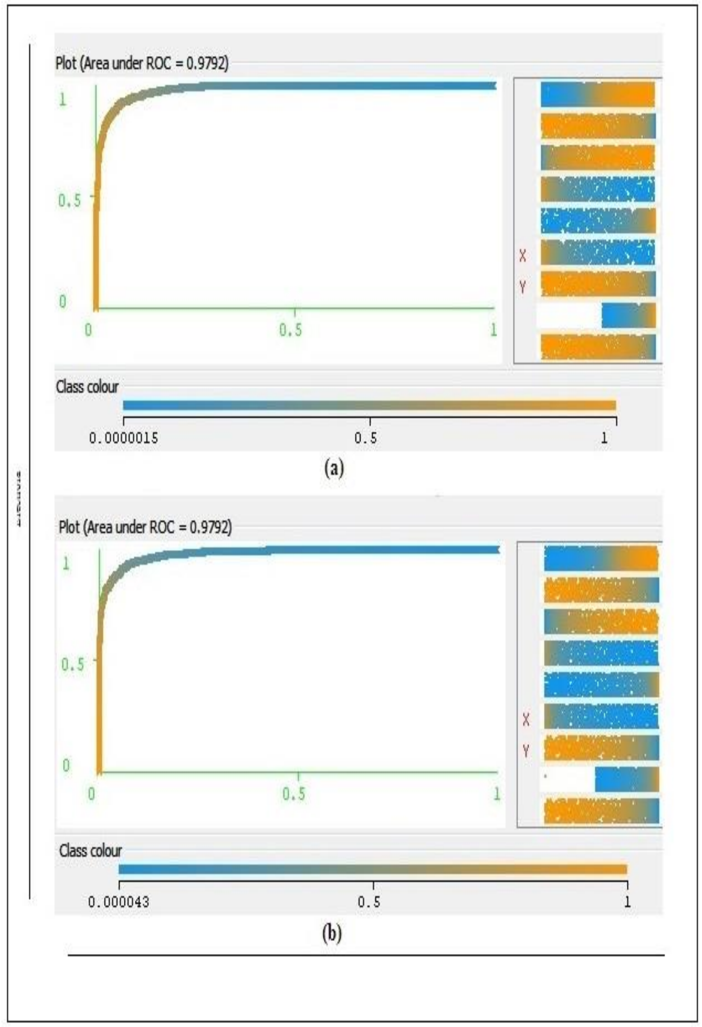

Fig. 7. ROC Curve removal on removal of F7, FC5, O2 (a) Threshold Value as 0 (b) Threshold value as 1

X. Conclusion

This is the first study to investigate the characteristics of Most Non Dominant feature from feature space they are less responsible to build the classification model, the MND set always gives concept which feature removal sufficiently reduce space and time requirement to build the classification model. This result is tested with EEG corpus to investigate eye state, either it is closed or open. Approximate 20% of time is saved by removal of these three most dominant features as compare to all attributes considered for classification.

TABLE III. RESULT ANALYSIS AFTER REMOVAL OF ATTRIBUTES FROM FEATURE SET FROM EEG DATA SET

Electrode Removal	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC	PRC	Accuracy %	Error %	Time Taken in sec.
Removal of F7+F3	**0.937**	**0.07**	**0.938**	**0.937**	**0.937**	**0.87**	**0.986**	**0.987**	**93.7103**	**6.2897**	**208.47**
Removal of F7+FC5	**0.927**	**0.08**	**0.929**	**0.927**	**0.927**	**0.85**	**0.983**	**0.984**	**92.74**	**7.26**	**194.78**
Removal of F7+O2	**0.939**	**0.066**	**0.94**	**0.939**	**0.939**	**0.88**	**0.985**	**0.986**	**93.9445**	**6.0555**	**202.76**
Removal of F3+O2	**0.947**	**0.057**	**0.948**	**0.947**	**0.947**	**0.89**	**0.989**	**0.989**	**94.714**	**5.286**	**209.24**
Removal of F7+FC5+O2	**0.92**	**0.089**	**0.921**	**0.92**	**0.919**	**0.84**	**0.979**	**0.98**	**91.9706**	**8.0294**	**184.06**

ACKNOWLEDGEMENT

This research is supported by the National Institute of Technology, Raipur and thanks to WEKA machine learning group as well as Rosler and Suendermann from Baden-Wuerttenberg cooperative state university (DHBW), Stuttgart, Germany for providing EEG corpus.

REFERENCES

[1] Jong-Min Lee, Dae-Jin Kim, In-Young Kim, Kwang-Suk Park, Sun I. Kim.: Detrended fluctuation analysis of EEG in sleep apnea using MIT/BIH polysomnography data, Computers in Biology and Medicine, Volume 32, Issue 1, Pages 37–47 (2002).

[2] M. V. M. Yeo, X. Li, K. Shen, and E. P. V.Wilder-Smith.: Can SVM is used for automatic EEG detection of drowsiness during car driving.Safety Science, vol. 47, no. 1, 115–124, (2009).

[3] Bulling A., Roggen D., and Troster G.: Wearable EOG goggles:eye-based interaction in everyday environments. ACM (2009).

[4] O. Rosler and D. Suendermann.: First step towards eye state prediction using EEG, Proceedings of the International Conference on Applied Informatics for Health and Life Sciences. Istanbul.Turkey(2013).

[5] T. Wang, S.U. Guan, K.L.Man and T.O. Ting.: EEG Eye State Identification Using Incremental Attribute Learning With Time-Series Classification. Hindawi Publishing Corporation Mathematical Problems in Engineering,vol- (2014).

[6] Mridu Sahu, N. K. Nagwani, Shrish Verma, Saransh Shirke,: An Incremental Feature Reordering (IFR) Algorithm to Classify Eye State Identification Using EEG, Volume 339, 2015, pp 803-811 edn., India: Information Systems Design and Intelligent Applications Advances in Intelligent Systems and Computing.

[7] Mridu Sahu, N.K. Nagwani, Shirish Verma, Saransh Shirke: Performance Evaluation of Different Classifier for Eye State Prediction using EEG Signal, Singapore: International Conference on Knowledge Engineering (ICKE 2015).

[8] G. Holmes, A. Donkin and I. H. Witten: Weka: A machine learning workbench, Proc Second Australia and New Zealand Conference on Intelligent Information Systems, Brisbane, Australia, Retrieved 2007-06-25, 1994.

[9] Sheng Chai, Jia Yang, and Yang Cheng: The Research of Improved Apriori Algorithm for Mining Association Rules. Proceedings of the Service Systems and Service Management, (2007).

[10] Ying Zhang , Almut Silja Hildebrand , Stephan Vogel: Distributed language modeling for N-best list re-ranking, Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, July 22-23,(2006) Sydney, Australia

[11] Liu, B., Hsu, W., Ma Y.: Integrating Classification and Association Rule Mining. KDD, (1998).

[12] Rakesh Agrawal , Ramakrishnan Srikant: Fast Algorithms for Mining Association Rules in Large Databases, Proceedings of the 20th International Conference on Very Large Data Bases, p.487-499, September 12-15, (1994).

[13] Sunita Sarawagi , Shiby Thomas , Rakesh Agrawal: Integrating association rule mining with relational database systems alternatives and implications, Proceedings of the international conference on Managementofdata,p.343-35,Seattle,Washington, USA(1998).

[14] Niedermeyer E. and da Silva F.L. Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Lippincot Williams & Wilkins. ISBN 0-7817-5126-8 (2004).

[15] Towle, Vernon L.; Bolaños, José; Suarez, Diane; Tan, Kim; Grzeszczuk, Robert; Levin, David N.; Cakmur, Raif; Frank, Samuel A.; Spire, Jean-Paul: The spatial location of EEG electrodes: Locating the best-fitting sphere relative to cortical anatomy. Electroencephalography and Clinical Neurophysiology 86 (1): 1–6. Doi:10.1016/0013-4694(93)90061-Y. PMID 7678386 (1993).

[16] Stephen J. Robert: Extreme value statistics for novelty detection in biomedical signal processing. In Proceedings of the 1st International Conference on Advances in Medical Signal and Information Processing. 166—172 (2002).

[17] V. Glandola , A. Banerjee and V. Kumar,: Anomaly Detection: A Survey, (2007).

[18] Pour, T. Gulrez, O. AlZoubi, G. Gargiulo, and R. Calvo.: Brain-Computer Interface: Next Generation Thought Controlled Distributed Video Game Development Platform in Proc. of the CIG, Perth, Australia (2008).

[19] Pham and D. Tran.: Emotion recognition using the emotivepoc device. Lecture Notes in Computer Science, vol. 7667(2012).

[20] O. Ossmy, O. Tam, R. Puzis, L. Rokach, O. Inbar, and Y. Elovici.: MindDesktop - Computer Accessibility for Severely Handicapped. Proc. of the ICEIS. Beijing, China (2011).

[21] J. van Erp, S. Reschke, M. Groojen, and A.-M. Brouwer: Brain Performance Enhancement for Military Operators in Proc. of the HFM. Sofia. Bulgaria. (2009).

[22] Yu L., Liu H.: Feature selection for high-dimensional data: a fast correlation-based filter solution. Mach Learn International Workshop Then Conf. 856(2003).

[23] N. Kwak and C.H. Choi.: Input Feature Selection by Mutual Information Based on Parzen Window. IEEE Trans. Pattern Analysis and Machine Intelligenc. vol. 24. no. 12,1667-1671,(2002).

[24] P. Langley.: Selection of Relevant Features in Machine Learning. Proc. AAAI Fall Symp. Relevance (1994).

[25] J. Jaeger, R. Sengupta, and W.L. Ruzzo.: Improved Gene Selection for Classification of Microarrays. Proc. Pacific Symp. Biocomputing, 53-64 (2003).

[26] R. Kohavi and G. John.: Wrapper for Feature Subset Selection. Artificial Intelligence. vol. 97. no. 1-2, 273-324 (1997).

[27] A. Webb.: "Statistical Pattern Recognition. Arnold" (1999).

[28] T. Cover and J. Thomas.: Elements of Information Theory. New York. Wiley. (1991).

[29] García, Salvador, et al.:A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability. Soft Computing 13.10 , 959-977(2009).

[30] Sokolova, Marina, and Guy Lapalme.: A systematic analysis of performance measures for classification tasks. Information Processing & Management 45.4, 427-437(2009).

[31] Demšar, Janez. : Statistical comparisons of classifiers over multiple data sets ,the Journal of Machine Learning Research 7 ,1-30(2006)

[32] T. Wang, S.U. Guan, K.L. Man and T.O. Ting.: EEG Eye State Identification Using Incremental Attribute Learning With Time-Series Classification. Hindawi Publishing Corporation Mathematical Problems in Engineering, vol- (2014).

[33] John G. leary, Leonad E Tigg.: K*: An Instance-based Lerner Using an Entropic Distance Measure, (1995).

[34] P.A. Est´evez, C. M. Held, C. A. Holzmann et al.: Polysomnographic pattern recognition for automated classification of sleep-waking states in infants. Medical and Biological Engineering and Computing, vol. 40, no. 1, 105–113 (2002).

[35] K. Polat and S. G¨unes: Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform .Applied Mathematics and Computation, vol. 187, no.2, 1017–1026 (2007).

[36] K. Sadatnezhad, R. Boostani, and A. Ghanizadeh.: Classification of BMD and ADHD patients using their EEG signals.Expert Systems with Applications,vol. 38, no. 3,1956–1963, (2011).

[37] N. Sulaiman, M. N. Taib, S. Lias, Z. H. Murat, S. A. M. Aris,and N. H. A. Hamid: Novel methods for stress features identification using EEG signals. International Journal of Simulation: Systems. Science and Technology. vol. 12, no. 1, 27–33 (2011).

[38] T. Nguyen, T. H. Nguyen, K. Q. D. Truong, and T. van Vo.: A mean threshold algorithm for human eye blinking detection using EEG. Proceedings of the 4th International Conference on the Development of Biomedical Engineering in Vietnam, 275–279.Ho Chi Minh City, Vietnam (2013).

[39] Thomas M. Cover, Joy A. Thomas: Elements of Information Theory Chapter 2 John Wiley & Sons, Inc. Print ISBN 0-471-06259-6 Online ISBN 0-471-20061-1 (1991).