International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Performance Analysis of Classifiers in Identifying NREM Sleep and Awaken Stages from EEG Signals

Harikumar Rajaguru*, Sanjai Khan N, Vinitha Krishnamoorthy

Abstract**: *Electroencephalogram is a medical procedure which helps in analyzing the activities of the brain through electrical signals. In this paper a simple classification technique of EEG signal into two stages as NREM sleep and awaken stages had been undertaken. Classifying these stages helps the physician to understand the patient's sleep disorder by knowing whether the person's brain is in NREM sleep or awaken stages. Physionet EEG signals are samples of 256 signals per second for 10 seconds duration is used in this work. Then the EEG samples properties are analyzed through various parameters like statistical features, entropy Pearson correlation coefficient, Power spectral density, scatter plots and Hilbert transform plots. The classification of NREM sleep and awaken stage is performed by the ten different classifiers broadly grouped into non linear and hybrid one. The classifiers used include Linear Regression, Non Linear Regression, Logistic Regression, Principal Component Analysis, Kernel Principal Component Analysis, Expectation Maximization, Compensatory Expectation Maximization, Expectation Maximization with Logistic Regression Compensatory Expectation Maximization with Logistic Regression, and Firefly. The performances of the classifiers are analyzed using regular parameters like sensitivity, accuracy, specificity, performance index. The highest accuracy of 95.575% is achieved with linear regression for awaken signal and an accuracy of 95.315% is achieved using kernel PCA for sleep signal.***

Keywords**: *EEG, NREM sleep, awaken, classifier, Performance.***

I. INTRODUCTION

Sleep disorder is one of the common disorders which affect many of the people around the world. It is a medical disorder in which the sleep pattern of the human or animal is changed and this may affect their health and day to day life. It disturbs the ability to sleep well in daily basis. It is common in both children and adult [1]. EEG stands for Electroencephalogram which is used to evaluate the brain's electrical activity. In other words, it is the recording of brain's activity. Analyzing the EEG signal of such sleep disorder patient and classifying those signals as NREM (Non Rapid Eye Movement) sleep and awaken helps the physician to identify the patient condition [2]. There are many techniques to classify them. Here we have taken the NREM sleep or awaken signal samples at 256 samples per second for 10 seconds using a 12 bit Analog to Digital convertor. So a total of 2560 signal samples are given as training as well as testing data [8]. Fig. 1 shows the work flow of the paper.

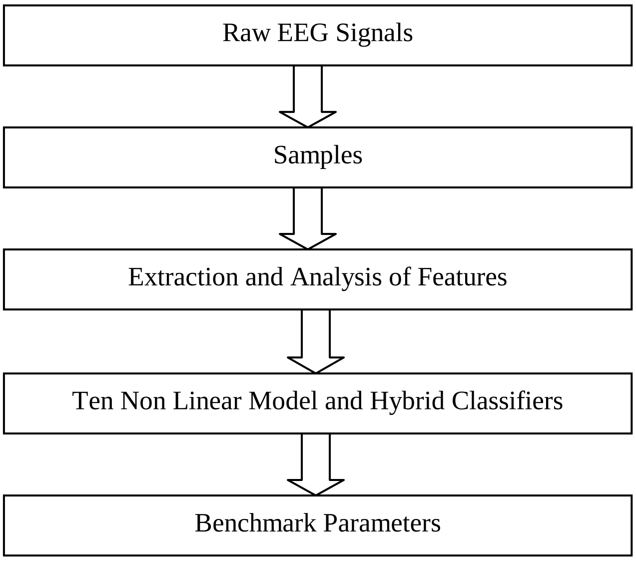

Figure. 1. Workflow of the Paper

These EEG signal samples are then send to the classifiers and a value is calculated. These classifier equations is implemented and final value is obtained. That value is compared with the target value which is set initially. Based on the deviation from the targeted value the performance of each classifier is analyzed by using performance factors such as sensitivity, specificity, performance index, accuracy and Mean square error. Thus, by the performance analysis of the classifiers we can identify the best classifiers. Fig. 2 depicts the plot of EEG samples for normal subject corrupted by motion artifacts in the Awaken stage. The EEG signal samples are preprocessed using noise removal algorithms like PCA and ICA techniques[4].

Manuscript received on March 15, 2020.
Revised Manuscript received on March 24, 2020.
Manuscript published on March 30, 2020.

* Correspondence Author
Harikumar Rajaguru*, Department of ECE, Bannari Amman Institute of Technology Sathyamangalam. harikumarrajaguru@gmail.com
Sanjai Khan N, , Department of ECE, Bannari Amman Institute of Technology Sathyamangalam.
Vinitha Krishnamoorthy, , Department of ECE, Bannari Amman Institute of Technology Sathyamangalam.

The motion artifacts free EEG samples are given as the input to the classifiers.

Fig. 3 depicts the EEG samples for NREM Sleep stage. It is observed from the figure 3 that the presence of unusual peaks at a definite interval indicates that the samples are associated with NREM sleep stage [3].

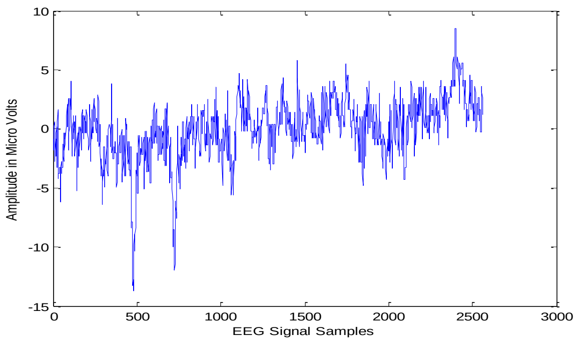

Figure. 2. EEG signal samples for Normal subject corrupted by motion artifacts

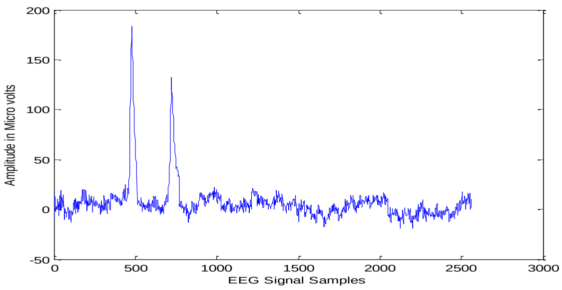

Figure. 3. EEG signal for NREM sleep subject

The histogram analysis explores the possibility of sparse components availability in the EEG samples. The histogram indicates whether the samples follow Gaussian or non Gaussian patterns. Fig. 4 demonstrates the histogram of Awaken EEG samples [9]. Figure 4 show that the histogram is a non linear one with wide base and containing more outlier values.

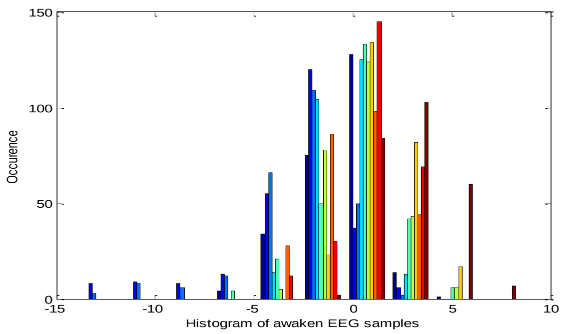

Figure. 4. Histogram of Awaken EEG samples

Fig. 5 shows the histograms of NREM sleep samples. From fig. 5 it is observed that the NREM sleep samples in histogram are skewed one and clustered at appoint. Therefore, the presence of nonlinearity expedite in the use of non linear and hybrid classifiers for the identification of NREM sleep and Awaken stages from EEG signals. The organization of the paper is as follows: In section 2, the materials and methods are discussed followed by the extracted feature analysis. In section 3, performances of ten different classifiers are analyzed. The results are discussed in the section 4. The results are concluded in the section 5.

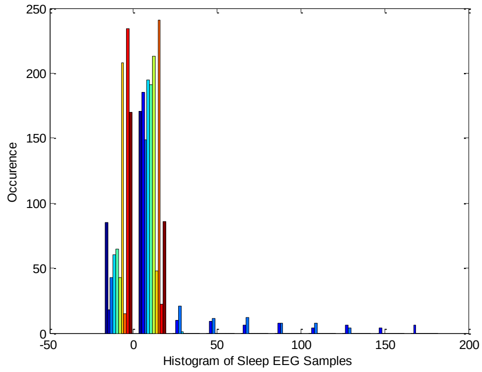

Figure. 5. Histogram of NREM Sleep EEG samples

II. MATERIALS AND METHODS

The physician dataset containing awakens and sleep signals undergo feature extraction. Feature extraction is done to extract the useful information from the image. Here mean, variance, correlation, kurtosis and skewness is used to extract the features. Mean is the average of the data in the dataset. Variance is the measure of difference between the random variable from the expected value. Correlation states how the variables are interrelated. Kurtosis states how the curve's tail of the distribution differs from the normal distribution. Skewness describes the asymmetry in the symmetrical curve [8]. By doing so the signal samples are reduced without information loss and the useful information is easily extracted. Table- I describes the definition of statistical features.

Table- I: Definition of Statistical Features

Features	Definition
1) *Mean*	$\mu = \frac{\Sigma x}{n}$ Where, x – observed data values, n - Complete number of values in that particular set of observations, μ - Arithmetic mean.
Variance	$\sigma^2 = \frac{\Sigma(x-\mu)^2}{n}$ Where, σ^2 – Variance, x- observed data values, n- Complete number of values in that particular set of observations.
Skewness	$g_1 = \frac{\sum_{i=1}^{n}(\mu_i - \bar{\mu})^3 / n}{\sigma^3}$ Where, n – No. of data points, $\bar{\mu}$ – mean and σ - standard deviation
2) *Kurtosis*	$K = \frac{\sum_{i=1}^{n}(\mu_i - \bar{\mu})^4 / n}{\sigma^4}$ Where, n – No. of data points, $\bar{\mu}$ - mean, and σ - standard deviation.

Retrieval Number: F9332038620/2020©BEIESP
DOI: 10.35940/ijrte.F9332.038620
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

To identify the distinct parameters of EEG signal statistical analysis is initiated.

The features such as mean, variance, skewness, kurtosis, Sample entropy, Permutation entropy, Correlation and Pearson correlation coefficient [9] are extracted from the EEG signals which are tabulated in Table- II.

Table- II: Average Parameters at Different features for NREM Sleep and Awaken case

Sl.no	Parameters	Sleep Stage	
		NREM	Awaken
1	Mean	6.61	-0.03
2	Variance	302.51	5.53
3	Skewness	20466.62	-11.34
4	Kurtosis	2529	186.14
5	Correlation	0.56	0.58
6	Pearson correlation Coefficient (PCC)	0.28	0.10
7	Permutation Entropy	1.44	1.44
8	Sample Entropy	5.42	3.82

It is observed from the Table- II that the statistical features of EEG sample among the classes are distinct but they are uncorrelated. The uncorrelated feature is predominant in the PCC as well as Correlation values. Further correlation analysis can be explored through the calculation of Canonical Correlation Analysis (CCA) [10] among the Awaken and NREM sleep samples and the same is tabulated in the Table- III.

Table- III Canonical Correlation Analysis

Sl.no	Awaken vs NREM Sleep
1	0.94
2	0.83
3	0.72
4	0.58
5	0.53
6	0.41
7	0.28
8	0.23
9	0.16
10	0.05
Average	**0.47732**

From Table- III it is observed that the awaken and NREM sleep stages are uncorrelated by means of low CCA values. The nonlinearity of the samples are further showcase in the Power Spectral Density (PSD) Plots for Awaken and NREM sleep stages. Fig. 6 and Fig. 7 demonstrate the PSD plots for Awaken and NREM EEG samples.

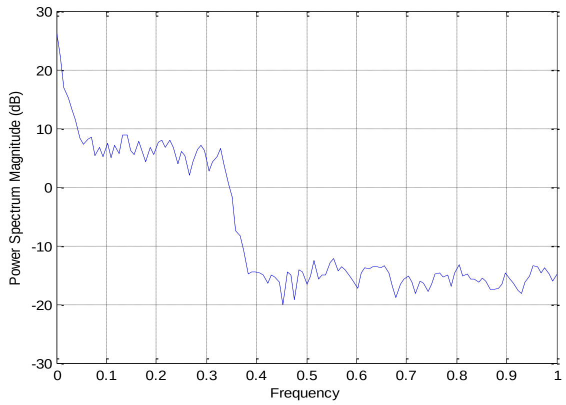

Figure. 6. Power Spectral density (PSD) plot for Awaken EEG Signals

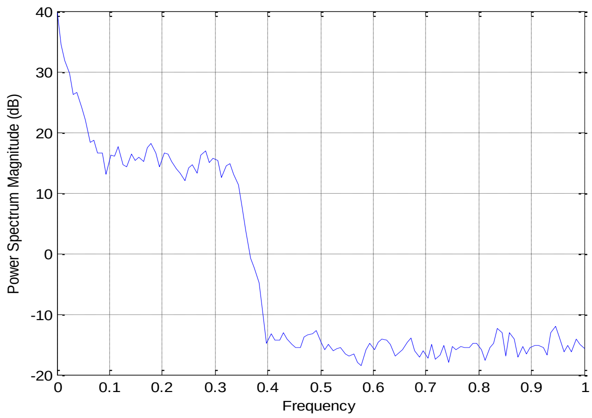

Figure. 7. Power Spectral density (PSD) plot for NREM Sleep EEG Signals

Fig. 8 displays the Hilbert Transform plot of Awaken and NREM Sleep EEG Signals. As from the figure 8 it is monitored that the Hilbert transform features are widely segregated among the classes.

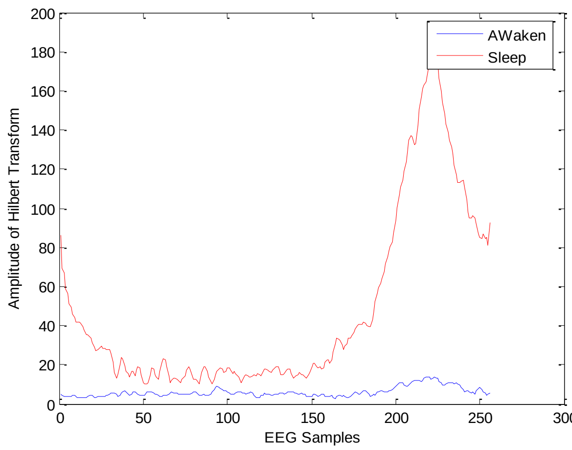

Figure. 8. Hilbert Transform of Awaken and NREM Sleep EEG Signals

Fig. 9 show the scatter plot for awakens and NREM sleep EEG signals. The clustered natures of EEG samples among the classes are well defined in this scatter plot.

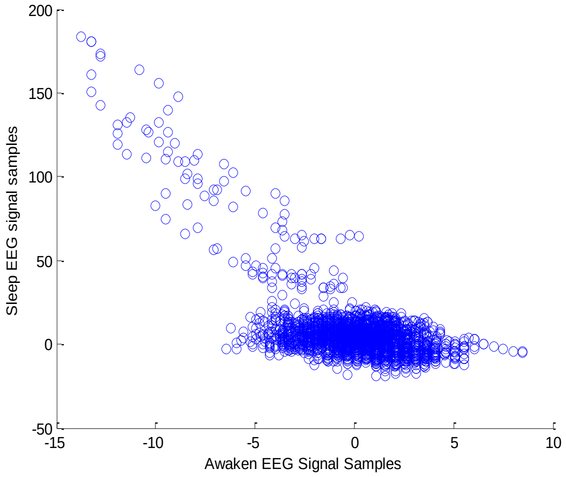

Figure. 9. Scatter plot for Awaken and NREM Sleep EEG Signals

The explanation about the non linear and hybrid Classifiers are discussed in the following section of the paper.

III. NON LINEAR AND HYBRID CLASSIFIERS FOR CLASSIFICATION OF NREM SLEEP AND AWAKEN STAGES

Classification is done using ten different classifiers by calculating a EEG sample value and applying it to the different classifiers. The classifiers used here are widely grouped as non linear and hybrid classifier.

A. Principal Component Analysis (PCA)

Principal Component Analysis is a dimensionality reduction tool which is used to convert a large set of data to small groups which will still contain the original information. By doing this, the classifier groups the similar data and based on the grouping the value for the sleep or awaken signal can be calculated. This method is highly used to analyze the large set of data and develop a predictive model [10]. Here the model is designed such a way that it predicts the sleep or awaken signal value based on residue value. Then the resultant value is being compared with the target value followed by performance factors calculation. Based on which the performance of the classifier can be analyzed.

B. Non Linear Regression

Nonlinear regression is one of the supervised learning methods in which the dependent variable or criterion variable depends upon a nonlinear function of one or more independent variables. Nonlinear regression is used where the curve should fit maximum number of points. That is, when there is many independent variables the linear line cannot satisfy the equation so we need a flexible curve to fit all the points. Here a residue value is calculated based on the target set. Then on applying the residue value and the sample dataset to the nonlinear equation we obtain a value and based on that the performance is analyzed.

C. Linear Regression

Linear regression is also a supervised learning method in which the dependent variable is linearly related to one or more independent variables. Linear regression may simple or multiple linear regressions where simple linear regression uses only one independent variable and multiple linear regressions uses more than one independent variable [5]. Here simple linear regression is used where a residue value is calculated based on the target value. The residue value is then applied to the linear regression equation. Based on the deviation from the target value, performance of the classifier is analyzed.

D. Logistic Regression

Logistic regression is also one of the supervised learning method. In this method, a logistic function is used to calculate the dependent binary variable though there are many complex points. This method used to identify the probability such as win or loss, pass or fail, etc. Here this classifier classifies as sleep or awaken. Then the performance is analyzed in the same way as the previous classifiers.

E. Expectation Maximization

Expectation Maximization (EM) classifier iteratively computes the expectation values (E step) and gives those values as the training dataset to the same model. Then based on the expectation value, the estimate value (M step) is calculated [6]. This is to maximize the estimate value more precisely. The reason why it works is that, the combination of E and M step minimizes the error and the error is nearly bounded to zero. Here the estimated value is calculated with the residue value and based on the deviation, the performance of the classifier is analyzed.

F. Compensatory EM

Compensatory EM is same as the Expectation Maximization but the only difference between the two is that, the weakness of one of the attributes is compensated for or by the strength on another [7]. The attributes are summed to discover the favorability of the attitude. Then the estimated value is calculated in the same way as the EM and the performance of the classifier is analyzed.

G. EM with Logistic Regression

Hybrid classifiers are nothing but the combination of one or more classifiers to increase the performance. One or more classifiers are cascaded to sum up the performance. Here the Expectation Maximization classifier is cascaded with the logistic regression model [6]. Using the Expectation maximization the parameters are estimated which is again given to the logistic regression model to get the precise value. Then the value obtained is compared to the targeted value and the performance factor for the classifier is calculated.

H. Compensatory EM with Logistic Regression

This is also one of the hybrid classifiers in which the compensatory EM is cascaded with logistic regreesion model. Here the parameters are estimated using compensatory EM and again given to the logistic regression model to get the optimal value [7]. The value obtained by doing so is then compared to the original target value to calculate the performance factors and analyze the performance of the classifier.

Retrieval Number: F9332038620/2020©BEIESP
DOI: 10.35940/ijrte.F9332.038620
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table- IV: Performance Analysis of Classifiers for NREM Sleep EEG Signal

Sl.no	Classifiers	PC	MC	FA	PI	Sensitivity	Specificity	Accuracy
1	PCA	59.64	0	40.36	32.15	59.64	100	79.82
2	Non Linear Regression	84.38	0	15.62	81.43	84.38	100	92.19
3	Linear Regression	79.69	0	20.31	78.42	79.71	100	89.85
4	Logistic Regression	57.09	0	42.90	24.80	57.09	100	78.54
5	EM	71.68	0	28.32	60.48	71.68	100	85.84
6	Compensatory EM	60.42	0	39.58	34.29	60.42	100	80.21
7	EM with Logistic Regression	66.21	0	33.79	48.91	66.21	100	83.10
8	Compensatory EM with Logistic Regression	63.8	0	36.19	43.22	63.8	100	81.9
9	Firefly	70.44	0	29.55	58.02	70.44	100	85.22
10	Kernel PCA	90.63	9.37	0	89.98	100	90.63	95.31

Table- V: Performance Analysis of Classifiers for awaken signal

Sl.no	Classifiers	PC	MC	FA	PI	Sensitivity	Specificity	Accuracy
1	PCA	73.43	0	26.56	63.76	73.43	100	86.71
2	Non Linear Regression	80.99	0	19.01	77.17	80.99	100	90.49
3	Linear Regression	91.15	0	8.85	90.78	91.15	100	95.57
4	Logistic Regression	73.43	0	26.56	63.76	73.43	100	86.71
5	EM	84.9	15.1	0	82.18	100	84.9	92.45
6	Compensatory EM	82.55	17.44	0	78.85	100	82.55	91.27
7	EM with Logistic Regression	60.94	39.06	0	35.72	100	60.94	80.47
8	compensatory EM with Logistic Regression	81.25	18.75	0	76.92	100	81.25	90.62
9	Firefly	81.77	0	18.22	77.69	81.77	100	90.88
10	Kernel PCA	81.77	18.22	0	77.69	100	81.77	90.88

I. Firefly

Firefly algorithm is a swarm intelligence algorithm that imitates the behavior of the firefly in nature.Classification based on firefly algorithm is processed by simulating the behavior of firefly in attracting other mates based on intensity and distances.Based on intensity, the training data set will be divided into firefly classes.Based on the distance, the appropriate firefly class will be found by the classifier[4]. Here based on the intensity and distance of the firefly class the appropriate class value will be calculated by the classifier and then the value is compared to the target value. Any deviation form the value is noted and the performance factors are calculated and after that the performance of the classifier is analyzed.

J. Kernel PCA

PCA is for linear dataset but if we use the same to the non linear dataset then the result obtained to the non linear dataset will not be obtimal or accurate since it is designed for linearly separable data.

So what is done in Kernal PCA is that, the non linear dataset is converted to linearly separable data. After that the same PCA method is used to predict the value of sleep or awaken signal[10]. Thus, these kind of classifiers holds good for non linear data. In case of linear data PCA is enough to develop a predictive model. Once the value is predicted the performance of the classifiers are analyzed in the same way of the previous cases.

IV. RESULTS AND DISCUSSIONS

Thus, the NREM sleep or awaken signals are taken as signal samples and then classified using different classifiers and the obtained value is compared to the original target values to get the performance factors such as Performance Index, Sensitivity, Specificity, Accuracy. Having these values we can estimate the best classifier.

$$PI = ((PC-MC-FA)/PC)$$

where, PI is the Performance Iindex, PC is the perfect classification, MC is the Missed classification, FA is the false alarm. From the perfect classification value, the missed classification and false alarm values are subtracted to get the performance index of the classifier. The sensitivity, specificity, accuracy are calculated using the following formulas as

$$\text{Specificity} = (PC / (PC + MC)) * 100$$

$$\text{Sensitivity} = (PC / (PC + FA)) * 100$$

$$\text{Accuracy} = (\text{Sensitivity} + \text{Specificity}) / 2$$

Specificity is calculated from perfect classification and the missed classification value. The Sensivity is calculated from perfect classification and the false alarm value. Accuarcy value is claculated from specificity and sensitivity values.

Table- IV shows the performance anlysis of classifiers for NREM sleep EEG signals. Higher accuracy of 95.31% is attained by the kernal PCA classifier where as PCA plugged in to the lower value of 79.82% of accuracy. Further all the nine classifiers except kernal PCA displays nil missed classification but with higher false alarm. This indicates that the classifiers are low threshols and least sen

V. CONCLUSION

This paper focused on the performance analysis of classifiers for NREM sleep and awaken stages of EEG signal.The highest accuracy of 95.575% is achieved with linear regression model for awaken signal and a accuracy of 95.315% is achieved using kernel PCA model for NREM sleep signal. Thus, the linear regression model holds well for awaken signal and the Kernel PCA models holds good for sleep signal. Further research will be in the direction of machine learning and deep learning algorithms for classification of sleep stages.

REFERENCES

1. M. Maeda, A. Takajyo, K. Inoue, K. Kumamaru, S. Matsuoka, "Time-frequency analysis of human sleep EEG and its application to feature extraction about biological rhythm'" in Proceedings of the SICE Annual Conference, Takamatsu, Japan, 17–20 September 2007, pp. 1939–1944.
2. L. J. Herrera, A. M. Mora, C. Fernandes, D. Migotina, A. Guillen, A. C. Rosa, "Symbolic representation of the EEG for sleep stage classification,", 11th International Conference on Intelligent Systems Design and Applications (ISDA), Nov 2011, pp.253-258.
3. S. Motamedi-Fakhr, M. Moshrefi-Torbati, M. Hill, C. M. Hill, P. R. White, "Signal processing techniques applied to human sleep EEG signals—A review", Biomed. Signal Process. Control 2014, 10, pp. 21–33.
4. J. C. Sackellares, D.S. Shiau, J. C. Principe et al., "Predictability analysis for an automated seizure prediction algorithm," Journal of Clinical Neurophysiology, vol. 23, no. 6, 2006, pp. 509–520.
5. A. Baykasoˇglu and F. B. Ozsoydan, "An improved firefly algorithm for solving dynamic multidimensional knapsack problems," Expert Systems with Applications, vol. 41, no. 8, 2014, pp. 3712–3725.
6. M. Vatankhah, M. Akbarzadeh-T, A Moghimi, "An intelligent system for diagnosing sleep stages using wavelet coefficients The 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, 2010, pp. 1-5.
7. E Estrada, H Nazeran, P Nava, K Behbehani, J Burk, E Lucas, "Itakura Distance: A Useful Similarity Measure between EEG and EOG Signals in Computer-aided Classification of Sleep Stages", Engineering in Medicine and Biology Society, Jan 2006, pp.1189-1192.
8. S. K. Prabhakar, H. Rajaguru and S. Lee, "A Framework for Schizophrenia EEG Signal Classification With Nature Inspired Optimization Algorithms", in IEEE Access, vol. 8, March 2020, pp. 39875-39897.
9. Sunil Kumar Prabhakar , Harikumar Rajaguru , Seong-Whan Lee, "Metaheuristic Based Dimensionality Reduction and Classification Analysis of PPG Signals for Interpreting Cardiovascular Disease", IEEE Access, vol .7 October 2019, pp165181-165206.
10. Harikumar Rajaguru, Sunil Kumar Prabhakar, "A Hybrid Classification Model Using Artificial Bee Colony with Particle Swarm Optimization and Minimum Relative Entropy as Post Classifier for Epilepsy Classification", Springer Lecture Notes on Computational Vision and Biomechanics, vol.28, February 2018, pp 593-603.

Retrieval Number: F9332038620/2020©BEIESP
DOI: 10.35940/ijrte.F9332.038620
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication