4th International Conference on Electrical Information and Communication Technology (EICT), 20-22 December 2019, Khulna, Bangladesh

Optimization of Sleep Stage Classification using Single-Channel EEG Signals

Md. Abdur Rahman
Electronics and Communication Engineering (ECE) Discipline
Khulna University
Khulna-9208, Bangladesh
abdurrahmanshohan15@gmail.com

Md. Abul Hossain
Electronics and Communication Engineering (ECE) Discipline
Khulna University
Khulna-9208, Bangladesh
abulhossain.torun@gmail.com

Md. Raihan Kabir
Electronics and Communication Engineering (ECE) Discipline
Khulna University
Khulna-9208, Bangladesh
raihanku320@gmail.com

Masrur Hossain Sani
Electronics and Communication Engineering (ECE) Discipline
Khulna University
Khulna-9208, Bangladesh
masrurhossainsani19@gmail.com

Abdullah-Al- Mamun
Electronics and Communication Engineering (ECE) Discipline
Khulna University
Khulna-9208, Bangladesh
mamun.eceku@gmail.com

Md. Abdul Awal*
Electronics and Communication Engineering (ECE) Discipline
Khulna University
Khulna-9208, Bangladesh
awalece04@yahoo.com ;
m.awal@ece.ku.ac.bd
*Corresponding Author

***Abstract*— Classification of various stages of sleep is mandatory for the diagnosis and treatment of sleep disorders. Manual scoring is a time-consuming and tedious task as well as it requires sleep specialists. Therefore, automatic sleep stage classification is necessary. In this paper, we have utilized state-of-the-art signal processing and machine learning techniques for sleep stage classification using single-channel EEG signal. Three cases of sleep classification have been done using support vector classifier (SVC), Decision tree (DT), Random forest (RF) and XGBoost (XGB). The features extracted from pre-procesed EEG have been applied to Spectral Regression dimensionality reduction technique to reduce the model complexity. The Bayesian Optimization (BO) technique is applied to optimize the hyper-parameters of the classifiers. Our proposed classification techniques provide the minimum error of 25.52%, 14.03%, and 4.93% for case I, case II and case III, respectively.**

Keywords*— *Sleep EEG, Machine Learning classification, Bayesian Optimization, Spectral Regression.

I. INTRODUCTION

"A sound sleep gives a man, a sound body and mind". Sufficient sleep is one of the basic needs of human beings, like other basic demands. For example, sufficient sleep makes our all-day-long activities enjoyable. In contrast, lack of sleep or insufficient sleep makes a man unsatisfied in mentally and physically with his apnea, Restless legs syndrome (RLS), Narcolepsy, circadian rhythm sleep disorders. In 2005-2006 a survey is performed in USA on 6,139 people over the age 16 [1]. The sleep disorder is higher for sleep works, makes him rude in his behavior and reduces his working efficiency. In addition, it interrupts his concentration on his daily routine and decreases his memory that causes sleep disorder such as insomnia, Sleep apnea (4.2%), insomnia (1.2%), RLS (0.4%) [1]. Note that, Sleep disorder can be defined as the problems of time, quality and the amount of sleep which interrupt our normal activities at daytime. There are many reasons for sleep disorder. The modernization of our societies and machinery lifestyle is the major causes for sleep disorder. Alcohol and cigarette are also responsible for changing the normal sleep patterns. Now, it is a burning question – "how can I understand that I am affected by sleep disorder?"

One possible way is monitoring his/her physiological data, such as electroencephalogram, heart rate variability, peripheral capillary oxygen saturation, Electrooculography and mark the data by an experienced clinician in order to provide his /her sleep quality or sleep disorder. Sleep quality can be measured by evaluating different stages of sleep. For example, Rechtschaffen and Kales (R&K) provide international guidelines for sleep staging. Currently, this sleep staging is done manually by an experienced technician which is very time consuming, especially for long recording. This situation becomes more crucial for developing like Bangladesh and its rural area. Where experienced technician is not available. To mitigate, signal processing and machine learning can be applied for automatic sleep scoring. Sleep staging can be done by single-channel and multi-channel signals. Though multi-channel signals provide unique performance, it is too much costly and it has connection complexity. In contrast, single-channel EEG is less complex and cheaper than multi-channel signal processing.

Previous studies e.g., a single channel EEG was used to complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) by Hasan et. al. [2], extracted linear prediction error energy, skewness, variance, kurtosis features and got 86.89% accuracy for six-class sleep scoring using partial least squares (PLS) algorithm and Adaptive Boosting (Ada-Boost). Similarly, [3] used Empirical Mode Decomposition (EMD) based features such as mean, variance, kurtosis, skewness and feed to Adaptive Boosting decision tree, Weak learn, support vector machine (SVC), least-square support vector machine (LSSVM), Discriminant Analysis (DA), support vector Discriminant Analysis (DA), Naïve Bayes, neural network, bootstrap aggregating (Bagging), and k-Nearest Neighbors (k-NN) classifiers to demonstrate automatic sleep scoring. [4] used entropy-based features and used dimensionality reduction and clustering techniques on two-channel EEG signals and obtained up to 80% accuracy. Different time-frequency based features extracted from short-time Fourier transform, Wavelet transform, Wigner-Ville Distribution (WVD) have also been utilized for sleep scoring [5]. Motivated by the above literature, we have extracted different features from

978-1-7281-6040-5/19/$31.00 ©2019 IEEE

frequency domain (F domain), time domain (T domain), time-frequency domain (T-F domain) and entropy-based features in this study.

In this paper, single-channel EEG signals are used for analyzing and classifying different sleep stages. In addition, we have also discussed various classification models and optimization techniques scoring the sleep stages by single-channel EEG signal.

II. MATERIALS AND METHODS

Figure 1 represents the flow diagram of the methodology. Firstly, the collected data is pre-processed to get clean EEG signals.

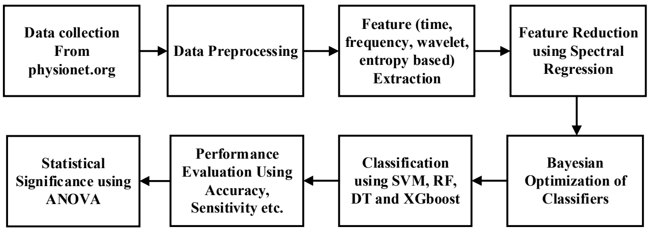

Fig. 1. Proposed methodology of this study.

This noise-free EEG data has been used for feature extraction and feature reduction respectively. After that state-of-the-art classifiers will be applied for classification. A Bayesian optimization algorithm will be applied to optimize the hyper-parameters of the classifiers. The reduced feature subset will be utilized to get optimized classification model. Finally, these optimized models will be tested on test dataset and different classification performance indices such as accuracy, sensitivity etc. will be used to evaluate the performance. In this study, the 10-fold cross-validation has also been used.

A. Data Collection and Preparation

Sleep EDF has versatile use as it is an open-source database in sleep scoring [6]. We have used expended sleep EDF database available at https://physionet.org/content/sleep-edfx/1.0.0/ [7]. We considered data from 17 Caucasians in this research and the age of the participants is between 21 and 35 years. They were abstaining from taken medicine. Sleep EDF database is divided into two parts i.e. sc* and st*, where sc* denotes sleep cassette and st* denotes sleep telemetry and they are presented in TABLE I. Dataset number starting with st* were taken in a hospital by a miniature telemetry system at night [2]. Only EEG signals from this database have been taken into account and they are located in Fpz-Cz and Pz-Oz positions. Previous research [1] has shown that EEG data from Fpz-Cz channel is less efficient than Pz-Oz channel for automatic detection of sleep staging. Therefore, we have used EEG channel collected from only Pz-Oz location. The sampling frequency is $100\ Hz$ [1]. An experienced technician marked the sleep data by following R&K criteria and they are denoted by AWAKE, REM, S1, S2, S3, S4, M, or 'U' classes to represent awake, Rapid Eye Movement, stage 1, stage 2, stage 3, stage 4, Movement and unscored class respectively. Among them, we have used six classes i.e. AWAKE, REM, S1, S2, S3, S4 class in this study. After that EEG signal is segmented on $30s$ basis for this study. Therefore, the epoch length is $30s$ or 3000 samples. For the classification purpose, we have used 3 cases (see TABLE I) to show how different classifiers classify these cases. A total of 16611 epochs have been used in this study (see TABLE II).

Case I: It consists of six classes i.e. AWAKE, REM, S1, S2, S3, S4.

Case II: It consists of three classes. Here S1, S2, S3 and S4 are combined together and denoted as Non-Rapid Eye Movement (NREM) class. AWAKE and REM are another two classes.

Case III: It consists of two classes. Here S1, S2, S3, S4 and REM are combined together and denoted as sleep stage class. AWAKE is another class.

TABLE I. CASES CONSIDERED FOR CLASSIFICATION IN THIS STUDY.

Case	Number of Classes	Constituent Sleep States	Dataset Number Used in this Study
I	6	AWAKE, REM, S1, S2, S3, S4	sc4112e0, sc4102e0, sc4012e0, sc4002e0, sc4001e0, sc4011e0, sc4032e0, sc4041e0, sc4042e0, sc4051e0, sc4052e0, st7241j0, st7212j0, st7132j0, st7121j0, st7052j0, st7022j0
II	3	AWAKE, REM, NREM (S1-S4)	
III	2	AWAKE, Sleep (REM & NREM)	

TABLE II. DESCRIPTION OF EPOCHS OF VARIOUS SLEEP STATES OF SLEEP-EDF DATABASE.

	S1	S2	S3	S4	REM	AWAKE	Total
Epochs No	1450	7843	1368	1026	3198	1726	16611

B. Data Pre-processing

The collected EEG signals have been filtered by using a third-order Butterworth filter within [0.5Hz – 30Hz] frequency band, as most of the sleep information lie in this frequency band.

C. Feature Extraction

The proposed EEG signals have been used to extract 125 features in time domain, frequency domain, time frequency (Wavelet) domain and entropy-based features.

TABLE III. DESCRIPTION OF THE FEATURES USED IN THIS STUDY.

Domain	Features name	Number of Features
Time	Total power, standard deviation, skewness, kurtosis, mean and standard deviation of envelop of Delta, Theta, Alpha and Beta signal	24
Range EEG	Mean, medium, lower margin, upper margin, width, standard deviation, coefficient of variation, Asymmetry of Delta, Theta, Alpha and Beta signal.	32
Frequency	Spectral power, spectral relative power, spectral flatness, spectral entropy, spectral edge frequency, spectral difference of Delta, Theta, Alpha and Beta signal.	21
Wavelet	Wavelet energy, coefficient of variance, standard deviation of wavelet coefficient.	44
Entropy	Approximate entropy, sample entropy, Hjorth parameters.	4

EEG signals have been decomposed into Delta (0.5 -4Hz), Theta (4 - 8Hz), Alpha (8 - 14Hz) and Beta (14 - 30Hz) bands and extracted different time domain features. In addition, different features from range domain EEG signals have also been extracted by following the methodology presented in [8]. The T domain EEG signals are transformed into F domain and several features extracted from this transformed signals presented in TABLE III. In addition, T-F domain features are extracted by decomposing the signal using fourth order Daubechies Wavelet transform with level=10. Different entropies play a significant role in EEG signal classification. TABLE III represents the entropies used in this study.

D. Feature Reduction

The Spectral Regression (SR), a dimensionality reduction method presented in [9] has been used in this study to reduce the dimensionality of the extracted features. The SR method mapped the feature set and reduce the feature dimension of $(class-1)$ features, where $class$ is the number of sleep classes. This method greatly reduced the dimension and therefore, simplifies the classification in terms of generating model and computation time.

E. Classification

The following four state-of-the-art classifiers are selected in this study as they are performed the best in most the classification problems [10]. A brief description of each classifiers is given bellow.

Random Forest: Many trees are generated by RF classifier from an input vector to classify a new object and the vector puts down each tree in the forest. Classification is given by each tree and each classification is considered as a 'vote' for that class. The forest selects the classification that have the major votes.

Support Vector Classifier: Support vector classifier (SVC) is a state-of-the-art machine learning classifier. It also called Support vector machine (SVM). The pre-lebelled training set $(\boldsymbol{x}_1 y_1), \ldots, (\boldsymbol{x}_n, y_n)$ where $y_i \in \{-1, +1\}$ for binary problem. For a test vector $\boldsymbol{x}$ can be classified using SVC by evaluating the following equation [11]:

$$d(\boldsymbol{x}) = \left(\sum_{i \in I_{sv}} \alpha_i y_i K(\boldsymbol{x}, \tilde{\boldsymbol{x}}_i) + b \right) \tag{1}$$

Where, sv represents support vector, $\alpha_i$is the weight. The kernel $K(,.,)$ of the SVC is used to map the input data into a higher dimensional feature space. This binary problem can be extended to multiclass problem using One-Vs-One and One-Vs-all. One-Vs-all is used in this study with Radial Basis Function (RBF) kernel.

Decision Tree: DT is a supervised machine learning algorithm. DT uses the feature values to classify a new instance. In DT, each node represents a feature of the instance that we have to be classified [12]. DT creates a training model that can use to predict class or by using decision rules inferred from training data. In DT, every internal node denotes a 'test' on an attribute, every branch denotes a result of the test and every leaf node denotes a class level.

XGBoost: XGBoost is one of the cutting edge algorithm to solve the classification, regression and ranking problems [13]. XGBoost is an ensemble machine learning algorithm which based on decision tree. A gradient boosting framework is used in XGBoost algorithm. In the XGBoost, the boosting trees are divided into two, one is regression trees and another is classification trees . The core function of the XGBoost algorithm is that optimized the value of the objective function [14]. A parallel tree boosting is provided by the XGBoost and it creates solution of many data science problem in an accurate and fast way. In boosting, hundreds of simple trees are combined with low accuracy to make a model that is more accurate. A new tree is produced by each iteration for the model. There the thousands of methods that create a new trees, a famous one is 'Gradient Boosting'. The gradient descent is used by 'Gradient Boosting to generate new trees and these trees based on all previous trees. It supervises the objective function toward the minimum direction [15].

F. Optimization the Hyper-parameters of Classifiers

Ultimately, the reduced feature set is used for classification by state-of-the-art classifiers and each classifier has some hyper-parameters as shown in TABLE IV. Classification performance such as classification accuracy, error etc. greatly depends on proper choice of the hyper-parameters. Therefore, optimization of the hyper-parameters is mandatory.

In this paper, we proposed to use a Bayesian Optimization (BO) algorithm. Bayesian Optimization (BO) is superior to grid search, random search and manual tuning. The BO memorizes the past evaluation and uses them formulating probabilistic Gaussian model of the objective function to find the optimal parameters. For details, see [16]. The cross-validation loss is used as the objective function of this study. To illustrate the support vector classifier with radial basis function (RBF) is chosen. It has two parameters that is (C and γ). For this classifier, C and γ parameters are chosen by using the BO technique for which the cross validation loss is the lowest. In this way, the hyper-parameters of the other classifiers such as RF, DT etc. can be optimized.

TABLE IV. CONTROLLING PARAMETERS OF DIFFERENT CLASSIFIERS.

Classifiers Name	Number of Controlling Parameters	Controlling Parameter's Name
RBF-SVC	2	Cost (C) and gamma (γ)
DT	3	Criterion (gini or entropy), max_features, max_depth
RF	4	Criterion (gini or entropy), n_estimators max_depth, max_features
XGBoost	6	learning_rate, n_estimators, n_jobs, max_depth, gamma, colsample_bytree

G. Performance Evaluation

To evaluate the classification performances, different performances measure indices such as accuracy, error, sensitivity, specificity and F1-score have been used in this study. These performance indices are calculated from the number of false positive (FP), false negative (FN), true positive (TP), true negative (TN). The equation of various performance indices are given bellow:

Accuracy (ACC)

$$ACC = (TP + TN)/(TP + FP + FN + TN) \tag{2}$$

Error = $1 - ACC$ (3)

Sensitivity (SNV) or true positive rate

$$SNV = {}^{TP}\!/_{(TP + FN)} \tag{4}$$

Specificity (SPC) or true negative rate (TNR)

$$SPC = {}^{TN}\!/_{(TN + FP)} \tag{5}$$

$F1\_$ Score is the harmonic mean of precision and sensitivity

$$F1\_\text{score} = {}^{2TP}\!/_{(2TP + FP + FN)} \tag{6}$$

H. Statistical Significance

To evaluate the statistical significance of different classifiers, one way is used to analysis of variance (ANOVA) and Box plot. The 10-fold cross validation technique has been used for this purpose. In 10-fold cross validation, one-fold will be used for testing and nine other folds are used for training and repeated ten times. This ten-fold cross validation provides 10 accuracy scores. ANOVA and Box plot have been applied on these scores. The p value produced by ANOVA will be used to test statistical significance. $p<=0.001$ is considered as statistical significance and $p>0.001$ is considered as statistical insignificant. This will show how a classifier is statistically different from other classifiers.

III. Results and Discussions

In this study, different results of three cases of sleep classification will be provided by following the methodology discussed above.

TABLE V. CLASSIFICATION PERFORMANCE OF DIFFERENT OPTIMIZED CLASSIFIERS.

Classification performance indexes (in %) For Case I					
Classifier Name	*ACC*	Error	F1_Score	*SNV*	*SPC*
Optimized SVC	74.48	25.52	62.50	62.52	93.98
Optimized DT	72.17	27.83	58.08	58.58	93.41
Optimized RF	73.23	26.77	58.64	59.50	93.52
Optimized XGB	73.57	26.43	59.97	61.13	93.73
For Case II					
Optimized SVC	85.81	14.19	79.55	79.39	90.13
Optimized DT	85.67	14.33	79.94	80.37	90.65
Optimized RF	85.97	14.03	79.96	79.94	90.49
Optimized XGB	85.51	14.49	79.10	79.25	90.23
For Case III					
Optimized SVC	95.07	4.93	97.27	98.17	68.28
Optimized DT	94.50	5.50	96.99	98.77	57.64
Optimized RF	95.05	4.95	97.25	97.79	71.37
Optimized XGB	89.66	10.34	94.54	100.0	0.19

For case I sleep, the Bayesian Optimized classifiers such as SVC, DT, RF, and XGB have been used and different performance indices such as accuracy, error, F1 score, sensitivity, specificity are shown in TABLE V. The Bayesian optimized SVC provided the highest accuracy of 74.48% among four classifiers. However, classification accuracy of different classifiers are close to each other when Bayesian optimized applied. The confusion matrix of the highest performing classifiers are shown in Fig 2(a) where we can visually see how different sleep classes are classified. In all three cases, similarly way confusion following the matrix of different classifiers can be presented.

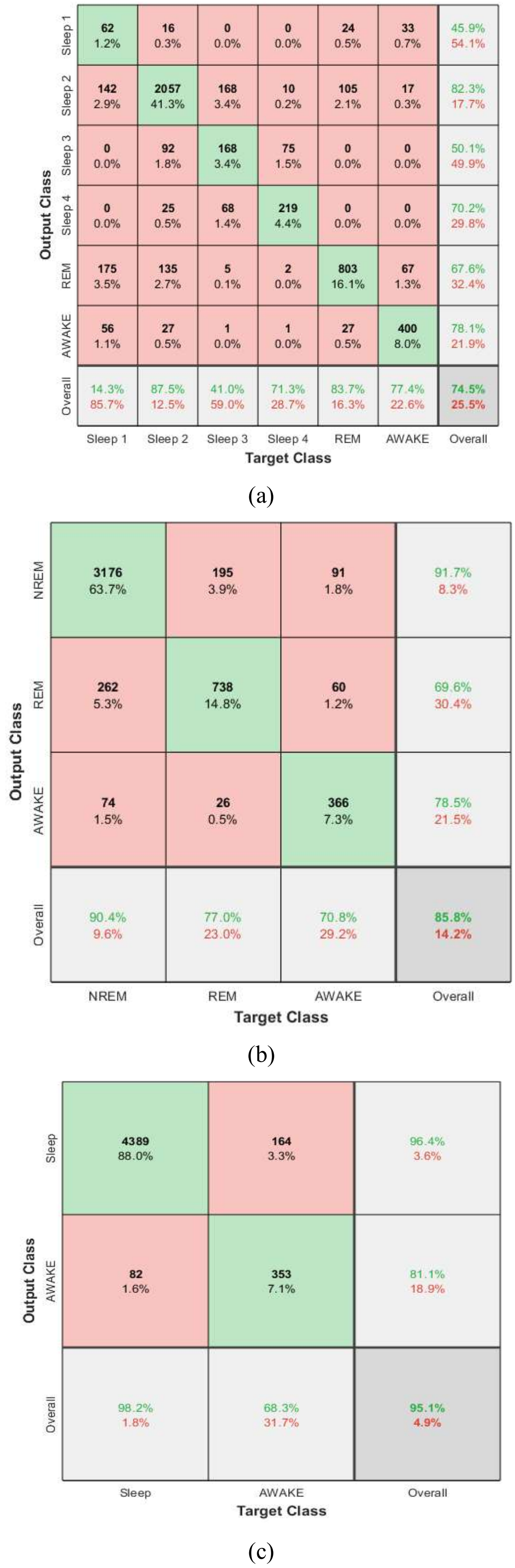

Fig. 2. Confusion matrix for (a) Case I, (b) Case II and (c) Case III sleep using optimized SVC.

For case II sleep study, different performance matrices for different optimized classifiers are shown in TABLE V. It can be seen that optimized RF provides the highest F1 score (79.96%). Due to space limitation, we have provided only the confusion matrix of SVC classifier. The confusion matrix of SVC classifier is presented in Fig 2(b). It can be seen that among 466 AWAKE plus 366 segments are correctly classified whereas 26 and 74 AWAKE class is misclassified to REM and NREM classes respectively. Other classes such as REM and NREM can be interpreted in the same way.

For case III sleep study, the performance matrices of different classifiers are shown in TABLE V. It can be seen that optimized SVC provides the highest accuracy of 95.07% that leaded to lowest error of 4.93%. The confusion matrix for SVC classifiers is shown in Fig 2(c). It can be seen that 4389 out of 4553 sleep EEG is correctly classified whereas 164 sleep class is misclassified to AWAKE class.

A. Box Plot and Statistical Significance

The accuracy of 10-fold cross validation has been used for Box plotting. Box plot for Case I is shown in Fig.3 and box plot for other sleep cases can be represented in the same way. One-way ANOVA has been utilized on 10-fold cross validation accuracy and p value has been calculated. The p value for 10-fold cross validation accuracy in case I, case II, case III is 0.001, 0.0023, and 1.73×10^{-28}, respectively which provides the statistically difference among classifiers.

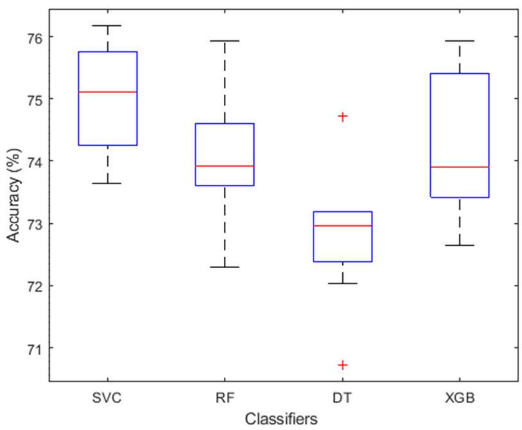

Fig. 3. Box plot for Case I sleep studies.

Regarding the comparison with other methods, Imtiaz et. al. showed how the classification results can be changed by using different set of training and test subject from the same database [17]. Therefore, we have not compared our results with other methods in terms of classification performance. The intention of this paper is to optimize use of machine learning classifiers for sleep staging with reduced number of features to make the model less complex and faster to execute. For comparision with our method (TABLE VI), Jian et. al used 151 features [18] and Hasan et. al. used 28 features [2], B.koley and D.Dey used 39 features [19] whereas our method used only 5, 3 and 1 feature for case I, case II and case III, respectively. It greatly reduces the computation time and simplifies the model for near-real time implementation.

TABLE VI. COMPARISION WITH OTHER STUDIES.

Authors name and Reference	**year**	**Number of features used**	**Features domain**
Jian et. al. [18]	2019.	151	Multi modal decomposition
Hasan et. al. [2]	2015	28	Complete ensemble empirical mode decomposition with adaptive noise(CEEMDAN)
B.koley and D.Dey [19]	2012	39	T domain, F domain and non-linear analysis
Hasan et.al. [3]	2015	28	Empirical mode decomposition
Proposed Method	2019	5, 3, 1 features for case I, case II and case III respectively	T domain, F domain, wavelet domain and non-linear analysis

IV. CONCLUSION

This paper represents a Bayesian Optimization framework for optimizing classifier's hyper-parameters and applied on sleep EEG data. The spectral Regression dimensionality reduction technique reduces the features dimension into only $(class - 1)$ features that speeds up the classification model. This technique simplifies the model as well as make very flexible and attractive for easy and near-real-time implementation in micro-controller-based devices to recognize different patterns such as sleep disorders for example, sleep apnea, fatigue and drowsiness detection. Our proposed method provides the accuracy as highest as of 74.48%, 85.97% and 95.07% for case I, case II and case III, respectively. Our future works are to add more data sets and compare our results with deep learning technique. Also examine the performance in noisy case of EEG. It is hoped that this research can help the researchers to implement a fully automatic and efficient sleep quality evaluation system.

ACKNOWLEDGMENT

The study is a part of a research funded by Khulna University Research Cell (KURC).

REFERENCES

[1] S. Ram, H. Seirawan, S. K. Kumar, and G. T. Clark, "Prevalence and impact of sleep disorders and sleep habits in the United States," *Sleep and Breathing,* vol. 14 (1), pp. 63-70, 2010.

[2] A. R. Hassan and M. I. H. Bhuiyan, "Automatic sleep stage classification," in *2015 2nd International Conference on Electrical Information and Communication Technologies (EICT)*, 2015, pp. 211-216.

[3] A. R. Hassan and M. I. H. Bhuiyan, "Automatic sleep scoring using statistical features in the EMD domain and ensemble methods," *Biocybernetics and Biomedical Engineering,* vol. 36 (1), pp. 248-255, 2016.

[4] J. Rodríguez-Sotelo, A. Osorio-Forero, A. Jiménez-Rodríguez, D. Cuesta-Frau, E. Cirugeda-Roldán, and D. Peluffo, "Automatic sleep stages classification using EEG entropy features and unsupervised pattern analysis techniques," *Entropy,* vol. 16 (12), pp. 6573-6589, 2014.

[5] K. Aboalayon, M. Faezipour, W. Almuhammadi, and S. Moslehpour, "Sleep stage classification using EEG signal analysis: a comprehensive survey and new investigation," *Entropy,* vol. 18 (9), p. 272, 2016.

[6] A. R. Hassan and A. Subasi, "A decision support system for automated identification of sleep stages from single-channel EEG signals," *Knowledge-Based Systems,* vol. 128 pp. 115-124, 2017.

[7] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. Kamphuisen, and J. J. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG," *IEEE Transactions on Biomedical Engineering,* vol. 47 (9), pp. 1185-1194, 2000.

[8] J. M. Toole and G. B. Boylan, "NEURAL: quantitative features for newborn EEG using Matlab," *arXiv preprint arXiv:1704.05694,* 2017.

[9] D. Cai, X. He, and J. Han, "SRDA: An efficient algorithm for large-scale discriminant analysis," *IEEE transactions on knowledge and data engineering,* vol. 20 (1), pp. 1-12, 2007.

[10] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?," *The Journal of Machine Learning Research,* vol. 15 (1), pp. 3133-3181, 2014.

[11] R. Ahmed, A. Temko, W. Marnane, G. Lightbody, and G. Boylan, "Grading hypoxic–ischemic encephalopathy severity in neonatal EEG using GMM supervectors and the support vector machine," *Clinical Neurophysiology,* vol. 127 (1), pp. 297-309, 2016.

[12] S. Dhaliwal, A.-A. Nahid, and R. Abbas, "Effective intrusion detection system using XGBoost," *Information,* vol. 9 (7), p. 149, 2018.

[13] T. Chen, T. He, M. Benesty, V. Khotilovich, and Y. Tang, "Xgboost: extreme gradient boosting," *R package version 0.4-2,* pp. 1-4, 2015.

[14] H. Zheng, J. Yuan, and L. Chen, "Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation," *Energies,* vol. 10 (8), p. 1168, 2017.

[15] L. Zhang and C. Zhan, "Machine learning in rock facies classification: an application of XGBoost," in *International Geophysical Conference, Qingdao, China, 17-20 April 2017*, 2017, pp. 1371-1374.

[16] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, 2012, pp. 2951-2959.

[17] S. A. Imtiaz and E. Rodriguez-Villegas, "Recommendations for performance assessment of automatic sleep staging algorithms," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 5044-5047.

[18] D. Jiang, Y.-n. Lu, M. Yu, and W. Yuanyuan, "Robust sleep stage classification with single-channel EEG signals using multimodal decomposition and HMM-based refinement," *Expert Systems with Applications,* vol. 121 pp. 188-203, 2019.

[19] B. Koley and D. Dey, "An ensemble system for automatic sleep stage classification using single channel EEG signal," *Computers in biology and medicine,* vol. 42 (12), pp. 1186-1195, 2012.