EEG Sleep Detection Algorithms a Comparative Study

Abu Saleh Musa Miah, Samiha Sahadat, Sonum Sabrin, Shagufta Tasnim, Anika Tabassum

Md. Mamunur Rashid

Abstract— **In this paper it represents a comparison between some machine learning algorithm which is applied to solve the sleep monitoring issues. Sleep detection requires electroencephalogram signal for differentiating purpose. There are some methods & models that are already perused & built with training & testing datasets like- single layer perception, multilayer perception, SVM (support vector machine),boosted tree method. The difference between these models is measured using the Cohen's index, the true positive & the true negative rate. Cross-validation technique is usually used to weigh the results of the models of monitoring sleep. The models successfully monitors sleep state reaching up to 94% and Cohen's index successfully reaching up to 0.69.The success rate shows the considerable assurance for future expansion & practices.**

Index Terms—EEG signal, Sleep stagine in EEG, feature extraction, EEG signal classification.

I. INTRODUCTION

Sleep plays a very important role in spending a quality life, maintain good mental & physical health. Problems like-sleep apnea, insomnia etc. are caused which are also known as sleeping disorder. Lack of sleep causes drowsiness. Now-a-days road crashes are happening and causing several fatalities. Drowsiness causes about 40,000 non-fatal damages and 1550 fatalities in the US alone annually [1]. The National Sleep Foundation in 2009 indicates that 1.9 million American drivers have had a car crash or 'near-miss' due to fatigue or drowsiness [2].

Abu Saleh Musa Miah, Department of Computer Science and Engineering, Bangladesh Army University of Science and Technology. Rangpur, Bangladesh, musa@baust.edu.bd.

Samiha Sahadat, Sonum Sabrin, Shagufta Tasnim Anika Tabassum, Department of Computer Science and Engineering, Bangladesh Army University of Science and Technology,Rangpur, Bangladesh.

Dr. Mamunur Rashid, Department of Computer Science and Engineering, Bangladesh Army University of Science and Technology, Bangladesh, mamunst2003@yahoo.com

The National Sleep Foundation indicates that some flights and train accidents have drowsiness as probable causes [3].In Bangladesh, according to World Bank statistics, annual fatality rate from road accidents is found to be 85.6 fatalities per 10,000 vehicles.[4]

First for detecting a driver's sleep stage the frequency spectra of brainwave signals are used as a differentiating factor. A better detection algorithm detects a drowsy driver accurately, does not cause false alarms. In practice, a monitoring model may detect drowsiness most of the time and the success rate could be of 100 % for sleep monitoring, but with this it will create some false alarms too, which would affect the drivers who are driving, behavior badly. When a wake stage is detected as a sleep stage the false alarm is triggered. While classification, a sleep stage is generally denoted as the first stage of sleep. Earlier experiments calculated 92.8 % internal sleep stage classification accuracy using multichannel electroencephalogram (EEG) signals [4] but this work did not classify wake stages then, which we would like to include in our developed models. The main objective of our study is to achieve better solutions for sleep detections that take fewer data as input parameters using only one channel. Moreover detecting sleep state is more important than having fewer false alarms.

The dataset used was obtained from the DREAMS project [5]. This dataset includes raw polysomnographic (PSG) signals. For training and testing the models, a data-mining plan was figured for transformation of the dataset and for appraisal of the models. The plan was expanded on the existing widely-used CRISP-DM model [6]. In this project, PSG data was gathered from sleeping patients. EEG signals, the backbone of our project was also taken from the patients. Automatically sleep stages and sleep disorders classification was done by DREAMS project team which was captured from different studies. It was educed that automatic classifications can be performed with the help of machine learning algorithms.

The previously done researches by various groups concludes that a good indicators of sleep and wake stages are EEG signals. In case of manual classification, there are two studies .A Manual of Standardized Terminology, Techniques, and Scoring System for Sleep Stages of Human Subjects done by US Department of Health, Education, and Welfare Public Health Service (1968). And the AASM Manual for the Scoring of Sleep and Associated Events. American Academy of Sleep Medicine (2007).Both of which relies on the frequency of the EEG signals.

In the contrary, an automatic detection of sleep stages using the EEG was done by Engineering in Medicine and Biology

Society. Proceedings of the 23rd Annual International Conference of the IEEE, 2, 1944-1947 (2001).

Again the automatic detection of the wake and stage 1 sleep stages using the EEG sub-epoch approach. It was done by Engineering in Medicine and Biology Society (EMBC) 2013, 35th Annual International Conference of the IEEE, 6401-6404 (2013).

In above two works it was exhibited that classification accuracy might be a bit advanced by using the complete frequency spectrum instead of few parameters.

In the project, the standard CRISP DM methodology was applied [6]. The goal of the paper is to built an automatic detection method and mechanism using data mining. The very first step is to understand the important physiological processes of sleep and their mathematical representation as signals. The following section provides a brief description of the brainwave signals along with their classification based on the frequency range. Then the data preparation phase is explained followed by an introduction to the classification models used in this project. In the third section, the experimental results for the different models are tabulated in terms of the true positive and true negative rates, also known as sensibility and specificity respectively. The paper ends with the conclusions drawn from our work and perspectives to continue research in this domain.

II. DATA SOURDCE AND CLASSIFICATION MODEL

A. *EEG signals*

Submit your manuscript electronically for review.

An electroencephalography (EEG) is recording of brain activity taken in brain signal on the surface of scalp.EEG signal is divided into bands and its depends on their frequency range:

- Beta wave- β-waves have frequencies between 13 and 30 and these waves are brainwaves. They are detecting full awareness & brain activity.
- Alpha wave- – α-waves have frequencies between 8 and 13 Hz.They are usually connected to relax states.
- Theta waves – θ-waves have frequencies between 4 to 8 Hz.They are generally linked to Non-Repaid Eye Moment(NREM) sleep.
- Delta waves– δ-waves are under 4 Hz in the frequency. They are commonly correspond to slow wave sleep states.

There are two standards to divide EEG data into sleep stages. The older Rechtschaffen and Kales (R&K) standard is determined of the following stages – Wake stage, REM stage, Sleep stage S1, Sleep stage S2, Sleep stage S3, and Sleep stage S4 [9]. In recent American Academy of Sleep Medicine (AASM) standard, the stages are classified as – Wake stage, REM stage, Sleep stage N1, Sleep stage N2, and Sleep stage N3 [10].In DREAMS dataset hold additional stages. There is one sleep stage movement exists in R&K classification. The additional sleep stages are identified as unknown sleep stages. The technology used to denote the sleep stages in defined as:

- 0 = Wake stage or REM (R&K and AASM)
- 1 = Sleep stage 1 (R&K and AASM) or Sleep stage 2 (R&K and AASM)
- 2 = Sleep stage 3 (R&K and AASM) or Sleep stage 4

(R&K) or unknown sleep stage (R&K)
- 3 = Sleep stage movement (R&K) or an unknown sleep stage (AASM)

The cause of data points which may not indicate as a wake stage or sleep stage S1/N1, are included when one model compute a sleep stage 1, the another model may compute a sleep stage 3, an unknown sleep stage, or a sleep stage movement

B. *Data Preparation*

The dataset to be used in present work was gathered as part of the DREAMS project in Belgium and build up 32-channel polygraph whole-night polysomnographic readings of 20 healthy subjects. At least three of the 32 channels were EEG channels. The frequency of collected data was 200 Hz & stored in standard European Data Format(EDF).we are choose the channel for the project was the channel CZ-A1,a central lobe channeled data are selected the first sleep and wake stages. Each EEG data point build up one hypnogram rating based on the R&R model, and one hypnogram rating based on the AASM model and 1000 raw EEG signal points similarly to a manual rating of a 5 second time window. For processing in the statistical software R were read into these data points. This is a very large volume of data(32924*1000) to be provided for a machine learning algorithm. Feature selection was performed to decrease the size. Each data point was first transformed as per the Fourier transform to obtain the frequency spectrum from 0 Hz to 100 Hz (half the sampling frequency) with an accuracy of 0.2 Hz (inverse of the length of the signal).The sleep manuals indicate that frequencies up to 30 Hz don't provide correct information for sleep stages and frequencies above 30.5 Hz was not allowed. The size of data was still large(32924*153). The frequency was averaged to the nearest integer frequency to decrease the dimension of the problem. The size of the data was 32924*32 after the data selection, transformation and feature selection. The final data was collected to the R platform, a statistical software package costly used for explore large data.

C. *Quantitative Indicators Used for the Models*

Quantitative Indicators is used to classify the model so it is important to describe. There are two models, first is corresponding to AASM scoring method and second is R&K scoring method. AASM scoring model and R&K scoring model should not be compared.

Various indicators are available for classification model. At first confusion matrix (table 1) is generated in our paper where TP means True Positives, TN means True Negatives, FN means False Negatives, FP means False Positives.

Table 1: Example of a Confusion Matrix

			Model's	Output
			0	1
Manual Rating		0	TN	FP
		1	FN	TP

False Positives: The value of false positive is measured through the number of sleep and wake states so it can play an important role to grade the models. The number of sleep and wake state in database is not equal so the following indicators are used to get a proper result along with Cohen's index:

- True Positive Rate (TPR):The value of TPR give a score of successful sleep detection rate and known as sensitivity of model.

$$TPR= TP / (TP+FN) \qquad (1)$$

It is linked to False Negative Rate(FNR) with TPR=1-FNR[12]

- True Negative Rate (TNR): Successful wake detection rate is calculated through the values of TNR and it is known as specificity of the model.

$$TNR=TN/(TN+FP) \qquad (2)$$
It is connected to False Positive Rate(FPR)
TNR=1-FPR[12]

- Another indicator is Cohen's index (k)-k which calculate the inter-rater agreement for categorical objects and it keeps the agreement (TP & TN) into accounts occurred by chance [13].

$$k = \frac{p_o - p_e}{1 - p_e} \qquad (3)$$

$$p_o = \frac{TP+TN}{(TP+TN+FP+FN)} \qquad (4)$$

$$p_e = \frac{(TP+FN)(TP+FP)}{(TP+TN+FP+FN)^2} + \frac{(TN+FN)(TN+FP)}{(TP+TN+FP+FN)^2} \qquad (5)$$

K=1 can be verified when the two raters are in an agreement. If k>0.6 then it indicates a good level of agreement.

In this model if TPR increase then TNR will reduce and vice versa. As TPR and TNR is necessary in this model so, it is important to study both. They are simple indicator and to grade this model k was used.

III. CLASSIFYING ALGORITHMS

In this section presents the classifying algorithm used in our experiments are-Single Layer Perceptron, Multi-layer Perceptron, Support vector machine, XGBoost, Multi-agent model. The accuracy of these models was checked to get the best model. The SLP measures the sum of the weighted input variables. A good reference model is formed to implement. MLP consists of layers of neurons and known as neural network. The SVM model can predict better better than MLP. It provides more flexibility because it can generate multiple support vectors. The XGBoost model is a collection of decision trees. It is the variation of gradient boosting model and minimizes the loss function. We propose two algorithms to multi-agent classifier. These are consists of all the above models by a majority vote and by treating the sleep and wake states on a more equal footing

A. *Singla- Layer Perception*

The ratios of the EEG band to execute classification of the sleep and wake state had been used by E. Malaekah and D. Cvetkovic. The classifying inference was tuned in their case. We did not use the EEG bands and the ratio. We used an algorithm which can measure the criterion without any human interference.

SLP is an algorithm which is used for binary classification of an input. It is a linear classifier. It can make its decisions by a scalar product. Scalar product is of input vectors with a set of weights. A perceptron is known as simplest neural network which is consist of one neuron. SLP algorithm can be shown as:

$$y(\vec{x}) = \begin{cases} 1 \ if \ \vec{\omega}.\overrightarrow{x+b} > 0 \\ 0 \ Oterwise \end{cases} \qquad (6)$$

Here ,
$\vec{x}$ is the input vector $\vec{\omega}$
is the weight vector b is the bias. The output is 0 or 1. It is measured by the condition which is provided by the scalar product. The bias changes the position of the decision line. Learning algorithm change the weight vector. Particle Swarm Optimization learning algorithm was used here. Having a set of candidate solution which is called particles , PSO measures a problem. And by rule for updating the position of these particles and velocities PSO move these particles in the feature space. These optimization techniques can be used to solve noisy and irregular problems because it does not make assumption of optimized parameters. PSO have some benefits of Mono Carlo techniques where various particles find large space of solutions. The particle is guided by a 'force' to the final solution through classical optimization techniques such as gradient decent, this is another advantage.

The position which is best for the particle attracts it. This happens to all the particles. There is a term known as inertia term for which particles are moved around the straight line. To guide particles to better solution, three forces work together in different areas. User chooses the factor multiplying the three factors, the force due to the global best position and particles' best position and the inertia term. These control the behavior and response of the system and also the swarming rate. The inertia term was multiplied by 1 in our implementation. The force due to particles best position was multiplied by a random number between 0 and 0.4. The force due to global best position was multiplied by random number between 0 and 0.7. To make the particles move to not stop swarming, the inertia term was kept as unit value. The particles were attracted to their best position because 0.4 and 0.7 were maximum value. But the contribution of these attractions did not exceed the contribution of the inertia term. A preference of each particle to explore position near the global best position was indicated by the higher average value of the attractive force. The particles' explored position near their best position was ensured by the non-zero value of the force. Based on the 'goodness' of the results, the final value was measured.

B. *Multi-Layer Perception*

The condition of linear detachable is not matched with the data that has been provided in our case, so non-linear classifiers were treated instead. Multi-Layer Perceptron (MLP) was figured out by its design and selected as it is a

renowned neural network model. As in fact for Neural networks it will must have at least two layers, probability of having some hidden layers. The previous experiments [15] which had been done with two and three hidden layers, final experiment gave maximum sleep and wake detection which had been rated 69 % and 90 % accordingly. This designed model decreased gradually the learning time for the neurons [15] and the number of connections which can be compared with designed pattern of three hidden layers. A neural network with two hidden layers 31-46-10-1 had been selected therefore depending on this result. The default algorithm was the learning algorithm of this project with weight back tracking resilient back propagation. The mean squared error, which was the error function of this project had been minimized. When a model starts to express the noise there is a risk of over fitting which is one of the main disadvantages of neural networks model. A predictive power finally would probably fail for the connection between input variables and output value. By using 16,462 training points 1,953 weights were required to be found. Poor predictions could thus occur for over fitting in neural network models. The performance of the neural network is decreased due to reduction of neurons in hidden layers on the other hand increasing neurons occurred over fitting therefore the training phase results were comparatively better than the testing phase results. By developing the risk of over fitting, the data parameters for neural network model could be described accurately. The neural network models performed well than previous at identifying the wake stages for the less number of sleep stages.

C. *Support Vector Machine*

For classification and regression analysis [16],[17] Support Vector Machines (SVMs) algorithms had been used which are called supervised learning. By establishing a hyperplane or a set of hyperplanes in a space which are of higher dimension than the feature space, data points were mapped in the SVM model of the feature space keeping themselves in a long distance. The distances between these hyperplanes were made to be extreme and for this two or more hyperplanes were chosen keeping no gap between them. It had been needed two types of SVM model, one was a simple SVM model and another was a multi-class SVM model. The hypnogram sleep stage rating was clarified by the first one, i.e. the output of a mathematical function of the SVM was expressed this value. For a given SVM input vector, the hypnogram value was allowed by the second one as a label. This model was linearly inseparable means more tough to clustering of data. Using the functionality provided in the package e1071 through the function tune.svm() on 10 % of the data tuning was performed for two parameters, gamma and cost, before starting the training of the SVMs. While cost was the cost of violating the constraints of the SVM model, the parameter gamma was internal to the application and was required for all the non-linear kernels. SVM application recommended the results of this tuning based on the 10 % of the data which marked a value of 0.1 for gamma, and a value of 1 for cost. In the final training of the models AASM and R&K classification methods were implemented and was achieved the same values. To compensate the influence of the higher number of wake states in the dataset of the multi-class variant of SVM (unlike in the default-SVM) it was given a weight of 0.2 and the sleep states were given a weight of 1.0.

D. *Extreme Gradient Boosted Model*

Supervised machine learning algorithm called Extreme Gradient Boosted Model (XGBoost) had been used for classification and regression analysis, comprises of an altogether of 'weak' prediction models, which is designed of decision trees. It had been popular to be many times which was faster than standard application and an effective application of the gradient boosting framework. For transfiguring the weak prediction models into a strong model the boosting process needed to be dealt with. For example, since their precision were slightly better than random classification, decision trees with only one or two levels would be considered weak, that was the reason why were more tough to overfitting. With the intention of creating a strong learner, a boosted model was created a group of these models. The weak learners were functions of the loss (error) functions in various applications where the group was trying to be minimized. Improving itself by summating the gradient of the loss function in the next iteration [18], [19] and consists of calculating the loss function at each iteration of the gradient boosting model. For better decision trees at each level the highest depth was set to 5 for this work. The number of iterations was reduced by this, which means for correction of the loss function 40 additional trees were required. By measuring the contribution of each tree by 0.15 comparing the default value of 0.3, the learning process was made more conservative (less overfitting). After running multiple tests on the data these values were selected and had been utilized for the AASM and R&K models

E. *Multi Agent Model*

Two MultiAgent (MA) models were developed for maximizing the sleep and wake detection rates. Mostly, in an independent classification model an MA model lies with many 'agents'. These MA models had been assembled from all the preceding algorithms for this work. A weighted MA model and a democratic MA model had been selected for implementing algorithms [20] for this project in various ways. The weighted MA model used to determine the output of a given input by weighing the outputs of each of the agents on the other hand the democratic MA model made to determine the outcome of a given input by a prioritized vote. As democratic MA model had been directly relying on its component models, it had not any training or testing phase, due to it. For analyzing the provided model on the training and testing data sets, the outputs from the agents had been directly used. AASM classification and R&K classification were the two models which were made. As concluded in the SLP model for the weighted MA model the training phase had been consisted of optimizing the weights using an error function, and the optimization had been achieved with the Particle Swarm Optimization (PSO). The sum of the FNR and FPR had been selected for the error function. This error function was not biased towards one of the sleep or wake states therefore the rates had been independent of the number of states. The sleep and wake detections were maximized due to minimizing this function synchronously (TPR = 1 - FNR and TNR = 1 - FPR). For definitive, the weights of the SLP with the exceptions that dimensions had been reduced to 5 and the bias was not required, the parameters of the PSO

algorithm had been indistinguishable to those. For each training set two models were made to work among them one is AASM classification and another one is R&K classification.

F. *Neural network classifier*

Using standard back-propagation algorithm a three-layer feed-forward artificial neural network which contains one hidden layer and one output layer was trained as it is presented in the figure 1 .

In input layer an input vector was adjoined, to which all the inputs are distinguished to each hidden layer unit. The input layers are multiplied by weight vectors of all the units. All the inputs are summed up together of each unit then generate a value. By using a non-linear activation function the reduced value is transformed and for this the common asymmetric sigmoid function is used. In the final layer the output is measured by multiplying the output vector of the hidden layer by the weights of each. By calculating even more activations and summations of each unit we might get the actual output of a neural network.
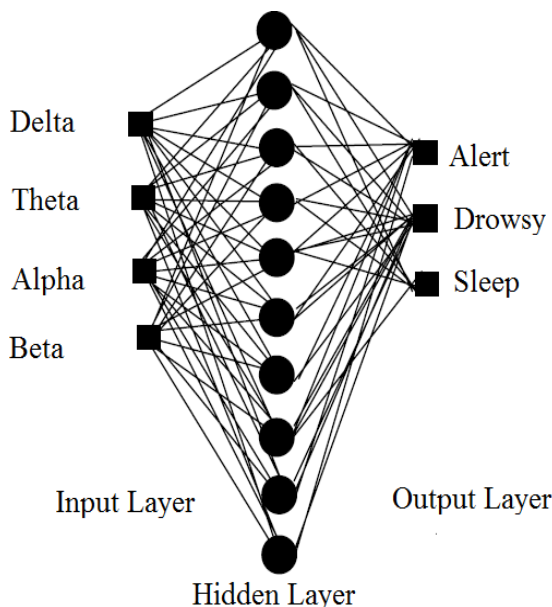

Fig. 1: Multi-layered neural network model

One of the most critical parts while designing a neural network is to designate the exact number of hidden layers of a network. There is no prior knowledge about the number of hidden layers unlike the input an output layers. A neural network that contains too mere hidden nodes or units would not be capable of distinguish the complex patterns that leads to nothing but a linear estimate of an actual trend. In contrary, if a network with too many hidden nodes or units, because of over-parameterization it will chase the noise of data and that would lead poor generation for the data that are still untrained. Again if the hidden layers increase due to the number of the layers the training will be so much time-consuming. The most effective procedure to measure the optimal number of the hidden layer of a network is- by trial and error (Basheer and Hajmeer, 2000; Fausett, 1994; Haykin, 1994). In this case, a neural network that contains one input layer, one output layer and one hidden layer

IV. EXPERIMENTAL RESULTS

To check the performance of the model under more general data, the data was partitioned randomly into training & testing sets for cross validation. 4-fold cross-validation was performed, where the data was partitioned into four equal sets. Previous experiments show same results to 2-fold cross-validation and there are more training and testing phases in 4-fold cross-validation, that's why we used it to obtain more results. We have shown our final results to compare among the models. For comparison, same data was given for training & testing sets in all learning algorithms.

Two scoring models (R&K and AASM) were built for each data point for each learning algorithm, e.g. MLP_Rnk and MLP_AASM. We need to compare all the AASM models together and all the R&K models together because there are two different standards. After breaking the data, they were divided into two sets, training-sets & test-sets then the data workspace was stored in R for further use & references.

Two outlier points (Outlier 1 and 2) are collected for having the same data point classified as sleep stages other than 0 or 1 in one of the two models, in the confusion matrix tables. For better analyze our results, these points were ignored. For comparing in terms of TPR, TNR and κ, the confusion matrix was interpreted again. For the 4-fold cross-validation studies in table II, these quantities are given.

In a general observation, it is proved that the multi-class SVM (R&K) performs the best which is followed by the multi-class SVM (AASM). The TPR for the multi-class SVM is about 94 % which is better than the previous results of 92.8 % classification accuracy observed by I.Zhovna and I.Shallom, where the information of the cross-correlation situated between the multichannel EEG signals was used. On the other hand, the TNR is 75 %, lower than the 77 % attained by E. Malaekah and D. Cvetkovic. This is the reason the multi-class SVM is a better choice for future study. As we can see, the multi-class SVM might be enhanced by giving more data points corresponding to sleep stages so that the FPR can be decreased.

Cohen's indices for different algorithms, we can see that the MA, XGBoost, MLP and SVM models are at least 0.60. But in our study, we give more emphasis on sleep detection, i.e. a TPR is more than 90 %. By using the multi-class SVM, we can acquire it. The XGBoost and the MA models have a TNR greater than TPR because the numbers of sleep states are lower than wake states.

From the tables, we suppose that the default SVM performs better than the SLP and MLP methods. Both default SVM models (AASM and R&K) have a TPR which is of over 70 %, same can be said for the TNR of the multi-class SVM, and similar to the TNR of the default SVM models. Thus the TPR and TNR values seem to be swapped when we compare the results of the multi-class SVM. This happens because of the larger number of the wake stages.

A value of κ > 0.60 shows a good contract between the different raters, and thus the SVM, XGBoost and MA models work better. The Cohen's indices indicates that for

both scoring methods the default SVM, XG Boost and MA models perform better with κ > 0.65.

The Receiver-Operator Characteristic (ROC) curves for all the models were plotted as shown in Fig. 1, to understand the trade-off between TPR and TNR. The FPR and TPR form the x−axes and y−axes of the graph and a guessing model will have TPR = FPR which is shown with the blue dotted-dashed diagonal in the figure. Points above the diagonal have TPR > FPR, it indicates good predictive performance. 'Perfect' classification model would be at the point (0,1) on the plot, i.e. FPR = 0 and TPR = 1.

We can see that the weighted MA (PSO) model works the best in terms of the trade-off, being close to the point (0,1), followed by the democratic variant based on the ROC curve. The error functions in the PSO algorithm acts on an equal footing for both sleep and wake stages. There is no training beyond for each of the individual agents in the democratic MA model. For this, the bias to the wake states exists because of the 'majority' of the models (MLP, XGB and SVM) being biased to the wake states, a result of the previously mentioned large number of wake states.

TABLE II:

COMPARISON OF RESULTS (4-FOLD CROSS-VALIDATION)

	AASM Criteria			R&K Criteria		
	TPR (%)	TNR (%)	Cohen's Index (k)	TPR (%)	TNR (%)	Cohen's Index (k)
SLP	72.3	76.9	0.47	73.0	77.1	0.46
MLP	69.0	88.2	0.58	70.5	89.3	0.61
Default-SVM	73.8	90.5	0.65	73.9	92.8	0.68
Multi-class SVM	94.0	67.0	0.53	94.0	74.9	0.60
XGBoost	74.5	90.6	0.66	76.2	82.2	0.69
Multi-Agent(Democratic)	77.2	87.8	0.65	79.1	90.1	0.69
Multi-Agent(PSO)	82.0	85.7	0.66	83.3	88.3	0.69

Detecting one stage might be more difficult and the 'best' model has to be chosen accordingly, depending on the requirement. The main motive of this project was to concentrate on sleep detection over wake detection. And for this, the multi-class SVM is a better model according to Cohen's index and the requirement of TPR > 90 %.

The performance of R&K models is better than AASM models. The R&K standard for sleep scoring was developed in 1968 on the other hand the AASM standard was developed in 2007. These differences might have affected the manual scoring of the used dataset. Inputs from a neuroscientist or more in vivo data based on a standard of counting drowsiness are required to choose between them. In addition, using a more balanced dataset with same numbers of wake and sleep stages may be helpful to enhance the methods further.

V. CONCULUSION

Mainly to reduce the rate of road accident and to warn the drowsy drivers, an automatic sleep monitoring device or system is needed and this is the main objective of this experimental work. To work with machine learning algorithms, we have used the DREAMS database [5] and also the input vector space was formed by using the Fourier transform of the EEG signal. Various experimental methods like- SLP, MLP, multi-class SVM, SVM, XG-BOOST etc. were performed in multiple types of models. To monitor the wake and sleep stage equally a new MA model was developed. With the help of ROC curve, it can be said that the MA model is the closest to a perfect classification. Also, the R&K scoring system that uses multi-class SVM model obtains the benchmarks of TPR 90 % and TNR 70 %. Again, from AASM scoring system the result is almost same. Due to the experimental benchmarks, the economic cost that was caused by road accidents was remarkably reduced.

In future, to ensure credibility and to confirm that the models won't fail when tested with new datasets, the models are needed to be examined using different datasets through future laboratory experiments. So, if the models are retrained for certain medical cases in those where the EEG signals might not be the same as an average human being, these models can be relatively compatible with multiple datasets efficiently.

Besides, it might be effective/operative to verify the experimental results from strict medical aspects, using EEG signal is preferable, to quantify drowsiness more than first sleep stages. Adding more data of a drowsy person like blood pressure or pulse, can be a help to improve the model as the physiological data decrease when a person is sleeping.

REFERENCES

[1] National Highway Traffic Safety Administration. Drowsy Driving and Automobile Crashes. http://www.nhtsa.gov/people/injury/drowsy driving1/Drowsy.html (Accessed 08 May, 2014).

[2] 1.9 Million Drivers Have Fatigue-Related Car Crashes or Near Misses Each Year. National Sleep Foundation (2009). http://www.sleepfoundation.org/media-center/press-release/ 19-million-drivers-have-fatigue-related-car-crashes-or -near-misses-each (Accessed 08 May, 2014).

[3] Crashes Where Fatigue Was a Contributing Factor. National Sleep Foundation (2012). http://sleepfoundation.org/sites/default/files/Crashes% 20Fatigue%20a%20Factor.pdf (Accessed 20 May, 2015).

[4] Zhovna I. and Shallom I. D.: Automatic detection and classification of sleep stages by multichannel EEG signal modelling. Engineering in Medicine and

Biology Society, 2008. 30th Annual International Conference of the IEEE, 2665-2668 (2008).

[5] Devuyst S., Dutoit T., Kerkhofs M.: DREAMS Project, The DREAMS Sleep Subjects Database. http://www.tcts.fpms.ac.be/~devuyst/Databases/ DatabaseSubjects/ (Accessed 16 April, 2014).

[6] KDnuggets Polls, Data Mining Methodology (2007). http://www. kdnuggets.com/polls/2007/data mining methodology.htm (Accessed 20 May, 2015).

[7] Van Hese P., Philips W., De Koninck J., Van de Walle R., Lemahieu I.: Automatic detection of sleep stages using the EEG. Engineering in Medicine and Biology Society. Proceedings of the 23rd Annual International Conference of the IEEE, 2, 1944-1947 (2001).

[8] Malaekah E. and Cvetkovic D.: Automatic detection of the wake and stage 1 sleep stages using the EEG sub-epoch approach. Engineering in Medicine and Biology Society (EMBC) 2013, 35th Annual International Conference of the IEEE, 6401-6404 (2013).

[9] Rechtschaffen A., Kales A.: A Manual of Standardized Terminology, Techniques, and Scoring System for Sleep Stages of Human Subjects. US Department of Health, Education, and Welfare Public Health Service (1968).

[10] Iber C., Ancoli-Israel S., Chesson Jr. A. L., Quan S. F.: The AASM Manual for the Scoring of Sleep and Associated Events. American Academy of Sleep Medicine (2007).

[11] Sucholeiki R.: Normal EEG Waveforms (2014). http://emedicine. medscape.com/article/1139332-overview#aw2aab6b3 (Accessed 14 April, 2014).

[12] Lalkhen A. G. and McCluskey A.: Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia, Critical Care & Pain, 8, 221223 (2008).

[13] Cohen J.: A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20, 37-46 (1960).

[14] Kennedy J. and Eberhart R.: Particle swarm optimization. Proceedings of IEEE International Conference on Neural Networks, 4, 1942 (1995).

[15] Pasieczna A. H.: An Approach to Driver Sleep Detection. Master Thesis Report, Wrocław University of Economics (2015).

[16] Boser B. E., Guyon I. M., Vapnik V.: A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory, 144-152 (1992).

[17] Cortes C., Vapnik V.: Support-Vector Networks. Machine Learning 20, 273-297 (1995).

[18] Mason L., Baxter J., Bartlett P. L., Frean M. R.: Boosting Algorithms as Gradient Descent. Advances in Neural Information Processing Systems 12, 512-518, MIT Press (2000).

[19] Friedman J. H.: Greedy Function Approximation: A Gradient Boosting Model, IMS 1999 Reitz Lecture (1999).

[20] Dietterich T. G.: Ensemble Methods in Machine Learning. Lecture Notes in Computer Science 1857, 1-15 (2000).

ABOUT THE AUTHORS

Abu Saleh Musa Miah is a Lecturer Department of Computer Science and Engineering in Bangladesh Army University of Science and Technology (BAUST), Saidpur, Bangladesh. He completed his B.Sc.Engg and M.Sc.Engg in Computer Science and Engineering from University of Rajshahi, Bangladesh. His research interests are artificial intelligence,Brain Computer Interfacing, EEG Signal, IoT, NLP, Satellite Image.

Samiha Sahadat B.Sc. Engineering Student of Bangladesh, Army University of Science and Technology, Rangpur, Bangladesh.

Sonum Sabrin B.Sc. Engineering Student of Bangladesh, Army University of Science and Technology.Rangpur, Bangladesh.

Shagupta Tasnim B.Sc. Engineering Student of Bangladesh, Army University of Science and Technology.Rangpur, Bangladesh.

Anika Tabassum B.Sc., Engineering Student of Bangladesh, Army University of Science and Technology.Rangpur, Bangladesh.

Dr. Mamunur Rashid is a Professor Department of Computer Science and Engineering in Bangladesh Army University of Science and Technology (BAUST), Saidpur, Bangladesh. He completed his B.Sc. Engineering and M.Sc. Engineering in Computer Science and Engineering from University of Rajshahi, Bangladesh. He completed his Ph.D. degree from Russia.