

An Investigation of Data Mining Based Automatic Sleep Stage Classification Techniques

Thakerng Wongsirichot, Nittida Elz, Supasit Kajkamhaeng, Wanchai Nupinit, and Narongrit Sridonthong

Abstract—Sleep quality is highly significant for the people's overall health. A standard diagnosis for sleep-related syndromes and illnesses is Polysomnography (PSG) or a sleep test in a controlled laboratory. However, PSG requires a sleep specialist to interpret bio-signals collected. It is a time consuming procedure. One of the fundamental step in the PSG is Sleep Stage Classification (SSC). In this study, we propose an investigation of Automatic Sleep Stage Classification (ASSC) using data mining techniques as an alternative to the PSG in order to reduce the time necessary for accurately diagnosing sleep quality. We studied 2,535 subjects' polysomnographic data with 14 channels of biomedical signals from the Sleep Heart Health Study (SHHS) Dataset. Subsequently, four data mining techniques including Decision Trees, Random Forests, Neural Network, and k-Nearest Neighbors were selected to compare the classification performances. The classification results in k-Nearest Neighbors achieved the greatest accuracy at 83.76%.

Index Terms—Automatic sleep stages classification, ASSC, sleep quality, data mining.

I. INTRODUCTION

Non-Communicable Diseases (NCDs) are chronic conditions or diseases that cannot be transmitted to others such as hypertension, hyperlipidemia, diabetes, cancers, heart disease, chronic respiratory diseases, etc. In recent years, people with NCDs are increasing dramatically. The causes of NCDs are primarily from various lifestyle and environmental factors including improper dietary intakes, uncommon sleep patterns, and unhealthy air conditions. According to World Health Organization (WHO), 48% of death worldwide in 2015, or 40 million over 56 million of global death, were categorized into NCDs and a large proportion of these deaths occurred before the age of 70. The magnitude of these diseases continues to rise, especially in low and middle income countries [1].

Sleep quality reflects the state of body and mind, as poor sleep quality and insufficient sleep relate to increasing risks of NCDs. Sleep problems can be in various forms and severity levels [2]. In medical practice, sleep quality and problems can be analyzed by Polysomnography (PSG) or a sleep test. PSG is considered to be the gold standard for

evaluating sleep stages and analyzing sleep quality.

PSG records at least 12 biological signal channels including Electrooculography (EOG), Electroencephalography (EEG), Electromyography (EMG), Electrocardiography (ECG), Airflow [3]. A sleep specialist interprets the collected PSG signals and analyzes individual sleep patterns and symptoms. The classification of sleep stages has been performed based on visual interpretations in every 30-second PSG epoch, according to the Rechtschaffen and Kales (R&K) sleep scoring recommendation. The R&K sleep scoring manual was developed to be a standard American Academy of Sleep Medicine (AASM). The sleep stages include wake (W), stage1 (S1), stage2 (S2), stage3 (S3), stage4 (S4), and Rapid Eye Movement (REM) [3].

II. THEORY

The CRISP-DM is a cross-industry process for data mining. The CRISP-DM methodology provides a structured approach to planning a data mining project. It is a robust and well-proven methodology. It contains the phases of a project, their respective tasks, and the relationships between these tasks. At this description level, it is not possible to identify all relationships could exist between any data mining tasks depending on the goals, the background, and the interest of the user—and most importantly—on the data. The life cycle of a data mining project consists of six phases, the sequence of the phases is not rigid. Moving back and forth between different phases is always required. This initial phase focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives. The data preparation phase covers all activities needed to construct the final dataset data that will be fed into the modeling tool(s) from the initial raw data. Data preparation tasks are likely to be performed multiple times and not in any prescribed order. Tasks include table, record, and attribute selection, as well as transformation and cleaning of data for modeling tools. In the modeling phase, various modeling techniques are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, going back to the data preparation phase is often necessary. At the evaluation phase in the project, you have built a model (or models) that appears to have high quality from a data analysis perspective. Before proceeding to final deployment of the model, it is important to thoroughly evaluate it and review the steps executed to create it, to be certain the model properly

Manuscript received December 13, 2018; revised June 1, 2019. This work was supported by The Innovative Information Technology for Health Science and Society (INTACT) Research Unit, Faculty of Science, Prince of Songkla University.

The authors are with the Department of Information and Communication Technology, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla, Thailand (e-mail: thakerng.w@psu.ac.th, nittida.n@psu.ac.th, supasit.k@psu.ac.th, wanchai.23.chai@gmail.com, iamnarongrit@gmail.com).

achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached. Creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it. It often involves applying “live” models within an organization’s decision making processes for example, real-time personalization of Web pages or repeated scoring of marketing databases. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise. In many cases, it is the customer, not the data analyst, who carries out the deployment steps. However, even if the analyst will carry out the deployment effort, it is important for the customer to understand up front what actions need to be carried out in order to actually make use of the created models [4].

Human sleep has been described as a succession of six recurring stages (including awaking). Sleep stage transition is characterized by abrupt changes in frequencies and amplitudes of the EEG signal. The first four sleep stages are defined as a “Non-Rapid Eye Movement” (NREM) sleep and the fifth stage is defined as a “Rapid Eye Movement” (REM) sleep. The NREM and REM sleep alternate in 90-100 minute cycles. A normal sleep pattern begins at about 80-90 minutes of NREM sleep, followed by an REM period for about 10-20 minutes. This cycle repeats about 4-6 times during the sleep [5].

The EEG signal is the most important signal in sleep stage classification but physicians more frequently make use of other biological signals such as EMG, EOG and ECG in manual sleep stage classification. At the wakefulness state the EEG pattern alternates between two main wave forms as the beta wave that has fast activity of 13-26 Hz and low voltage of 10-30 μ V, the alpha wave that has higher voltage of 20-40 μ V and slower activity of 8-12 Hz. The NREM sleep occurs for 75-80% of total sleep time and it is characterized by low frequency and high voltage wave activity that correspond to increasing depths of sleep. According to the American Academy of Sleep Medicine (AASM) the NREM sleep can be divided into four separate stages (stage 1 to stage 4). The stage 1 can be defines as a gateway state between the awake state and sleep state, the duration is about 5-10 minutes, This stage is characterized by relative low EEG voltage and slow movements of eye rolling, the alpha waves (8-13 Hz) had seen in the awake state and are replaced by theta waves (4-7 Hz). The stage 2 takes approximately 45-55% of the total sleep, it is characterized by a lack of eye movement, sleep spindles and K-complexes (A sleep spindles is a burst of brain activity visible on EEG, it consists of 11-15 Hz waves that occur for 0.5-1.5 seconds. A K-complexes is a sudden, brief, high amplitude waveform of EEG. It consists of a brief high-voltage peak, and lasts for longer than 0.5 seconds). The stage 3 refers to a deep sleep and happens for 35-45 minutes after falling asleep, it is characterized by 20-40% of delta (slow) wave and high amplitude (>75 μ V). Additionally, a K-complex and sleep spindle can appear at this stage. The stage 4 refers to a very

deep sleep, it is very similar to stage 3 in some cases and it is characterized by delta waves. The REM sleep refers to dream occurs, therefore a burst of prominent rapid eye movement appears in EEG, a mixed frequency and low voltage with occasional bursts of saw-tooth waves [6].

III. BACKGROUND

Data mining techniques can be categorized by purposes including association rule mining prediction, classification, clustering, and discovering new and useful patterns in particular problems. Problems usually comprise of large datasets with some degree of complexities. In classification problems, it is usually supervised learning techniques that classify data items into predefined class labels. Data mining projects are generally followed steps in the cross-industry standard process so called CRISP- DM. It consists of six major phases including Application Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment [4].

In CRISP-DM modeling step, a single or multiple data mining techniques are selected. There are a number of well-known data mining techniques such as Decision Tree (DT), k-Nearest Neighbor (k-NN), Neural Network (NN), Random Forest (RF), etc. DT is one of the most commonly used in classification problems. A DT is created by recursively selecting the best attributes to split the data and expanding the leaf nodes of the tree until a predefined stopping criteria is met. The splitting process is based on a gain ratio calculation, which automatically give weights to each of the studied attributes. The k-Nearest Neighbors (k-NN) is a non- parametric method created by assigning weights to the contributions of the neighbors. Nearer neighbors contribute more to the average than more distant ones. The Neural Network (NN) is typically organized in layers, created by using a number of interconnected call nodes. An activation function is identified and interacts with one or more hidden layers where the actual processing is done via a system of weighted connections, which then link to an output layer with the answer. Random Forest (RF) is an ensemble learning method used in various classifications. It is constructed by multiple decision trees at the training time.

Over the years, a number of automatic sleep stage classification (ASSC) techniques have been presented. A novel audio-based method was proposed to estimate four sleep stages as wake, light sleep, deep sleep and Rapid Eye Movement (REM) from nocturnal audio signals. The method consisted of three steps, breathe/snore detection based on Adaptive Vertical Box (AV-Box), nonlinear feature and sleep pattern feature extraction, and sleep stage classification utilizing a decision tree classifier. A seven hours long nocturnal audio signals were recorded for one healthy male subject with PSG in a sleep laboratory at Nanjing University of Science and Technology. Performance evaluations of six types of decision tree classifiers was conducted using the k-fold cross validation technique. The average prediction accuracy is 74.3% [7]. Another research work was observing a sleep stage classification from EEG. It investigated ASSC from EEG using complexity science methods that applied fuzzy entropy and permutation entropy as kernels of multi-scale entropy analysis. A combination of signals from

three sources including a current epoch, the preceding of the current epoch, and the following of the current epoch were selected for analysis. Combining the entropy and spectral edge frequency features extracted from one EEG channel, a multi-class support vector machine (SVM) was able to classify 93.8% of the 5 sleep stages for 8 subjects in the Sleep EDF database. The sensitivity of S1 stage was 49.1% [8]. However, these research works were limited by a small number of studied subjects. A further expansion of studied subjects will ideally conclude the work. Another research of sleep stages classification using Electrooculogram (EOG) signals by analyzed in Discrete Wavelet Transform (DWT) domain employing various statistical features such as Spectral Entropy, Moment-based Measures, Refined Composite Multiscale Dispersion Entropy (RCMDE) and Autoregressive (AR) Model Coefficients. The discriminating ability of the features is studied using the One Way Analysis of Variance (ANOVA) and box plots. A feature reduction algorithm based on Neighborhood Component Analysis is used to reduce the model complexity and select the features with highest discriminating abilities. Random Under-Sampling Boosting (RUSBoost), Random Forest (RF) and Support Vector Machine (SVM) are employed to classify various sleep stages for 2–6 stage classification problem. Performance of the proposed method is studied using three publicly available databases, the Sleep-EDF, Sleep-EDFX and ISRUC-Sleep databases consisting of 8, 20 and 10 subjects respectively. The proposed method outperforms the state-of-the-art EOG based techniques in accuracy. In addition, its performance is shown to be on par or better than those of various single channel EEG based methods. An important limitation of existing sleep detection methods is the low accuracy of the S1 sleep stage classification for which the proposed method using the RUSBoost classifier gives a superior accuracy as compared to those of EOG and EEG based techniques. SVM and RF have overall higher accuracy, RUSBoost performs consistently better to classify S1 sleep stage with satisfactory accuracy while giving comparable overall accuracy value. It suggests that the classification accuracy of most prior sleep stage classification methods is heavily influenced by the imbalance of the dataset. About 53.04% and 68% of the data of Sleep-EDF and Sleep-EDFX belong to the Awake (W) class. This is important since, for example for 6-class classification, a significant high accuracy is achieved for the Awake class

where the accuracy of the S1 class is relatively low. Thus, the overall accuracy is basically determined by the accuracy of the Awake class. In our experiment, we have used the ISRUC-Sleep database which is more balanced. Notice that with a relatively balanced database, SVM and RF still performs poorly to classify the S1 sleep stage. However, RUSBoost consistently produces a satisfactory performance [9].

In this work, we use multi biological signal for 14 channel such as EOG, EEG, EMG, ECG, PR etc. it decomposed into time-frequency domain components using various statistical features are then extracted from the decomposed signals such as maximum, minimum, standard division, average and kurtosis. Finally, classification of the features is carried out using three widely known machine learning algorithms as Random forest (RF), Support Vector Machine (SVM). We used publicly available dataset were Sleep Heart Health Study (SHHS) consisting of 2,535 patients respectively are used to study the accuracy of detection of individual sleep stages as well as 6 class sleep stage classification accuracy.

IV. MATERIAL AND METHOD

A. Dataset

A PSG recording dataset, named the Sleep Heart Health Study (SHHS) Dataset, was used in this study. It was collected using a multi-center cohort study of the National Heart Lung & Blood Institute to determine the cardiovascular and other consequences of sleep disorders. Sleep-related breathing data associated with increased risks of coronary heart disease, stroke, all-cause mortality, and hypertension were collected. Data were obtained from 2,535 of the participants aged 40 years and older were enrolled between January 2001- June 2003 [5]. All recordings included 14 channels of biomedical signals are shown in Table I.

TABLE I: 14 CHANNELS BIOMEDICAL SIGNALS OF IN SHHS DATASET

Channel	Sampling Rate (Hz)	Channel	Sampling Rate (Hz)
SaO2	1	Abdo RES	10
PR	1	EOG(L)	50
Position	1	EOG(R)	50
Light	1	EEG	125
OXSTAT	1	EEG (sec)	125
Airflow	10	ECG	250
Thor RES	10	EMG	125

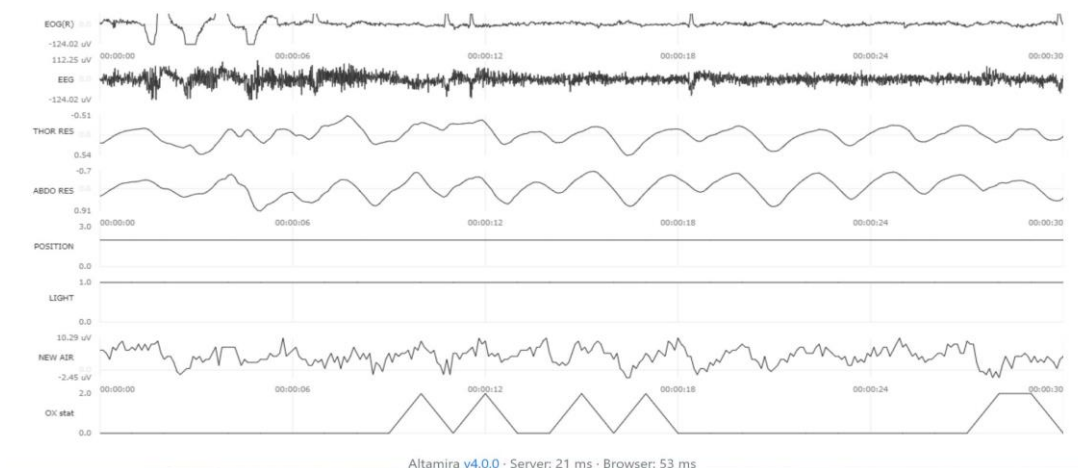


Fig. 1. Example data from EDF file that contain 14 channel of biomedical signals.



Fig. 2. Overall data from XML annotation file that contain 6 sleep stages.

Sampling Rate	Sample
1 Hz	15,660 – 50,730
10 Hz	156,600 – 507,300
50 Hz	783,000 – 2,536,500
125 Hz	1,957,500 – 6,341,250
250 Hz	3,607,500 – 12,682,500

In this dataset, sleep recordings were scored using the Rechtschaffen and Kales (R&K) rules with 30-second intervals. Each epoch has a single sleep stage called hypnogram, defined as WAKE, NREM1, NREM2, NREM3, NREM4, and REM, by the R&K criteria [7], [10]. In addition, biomedical signals were recorded in European Data Format (EDF), which is the standard file format designed for the exchange and storage of medical time series and a flexible format for the exchange and storage of multichannel biological and physical signals. In addition, hypnogram data is stored in an Extensible Markup Language (XML) file. Each dataset consists of different samples in subjects, as shown in Table II. The data from EDF contain 14 biomedical signals such as EOG, EEG, EMG, ECG, PR etc. that shown in Fig. 1 and The data from xml contain hypnogram data of sleep stages in total sleep time that shown in Fig. 2.

B. Data Preprocessing

Each channel consists of different N samples from 14 channel biomedical signals. Each of the extracted records was calculated in various epochs, 30 seconds, 15 seconds, 10 seconds and 5 seconds according to Equation 1.

$$\text{Interval of signals} = \frac{\text{Sample}_i}{\text{epoch} \times \text{SamplingRate}_i} \quad (1)$$

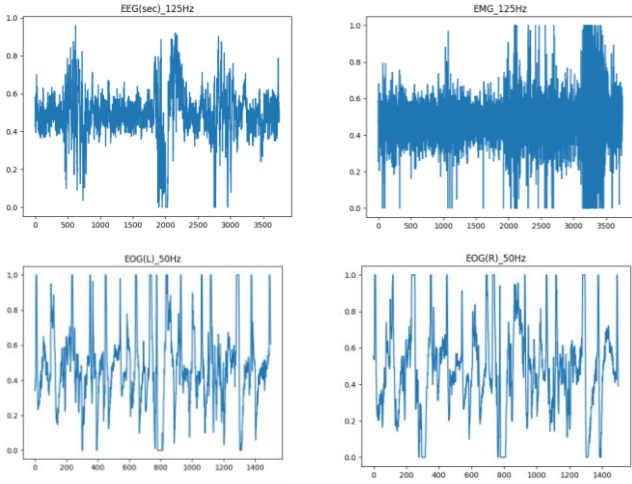


Fig. 3. Example when extracted biomedical signal from Equation 1 with epoch 30 seconds in 1 interval.

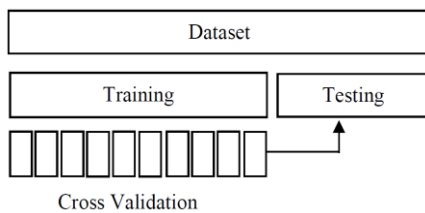


Fig. 4. K-fold cross validation.

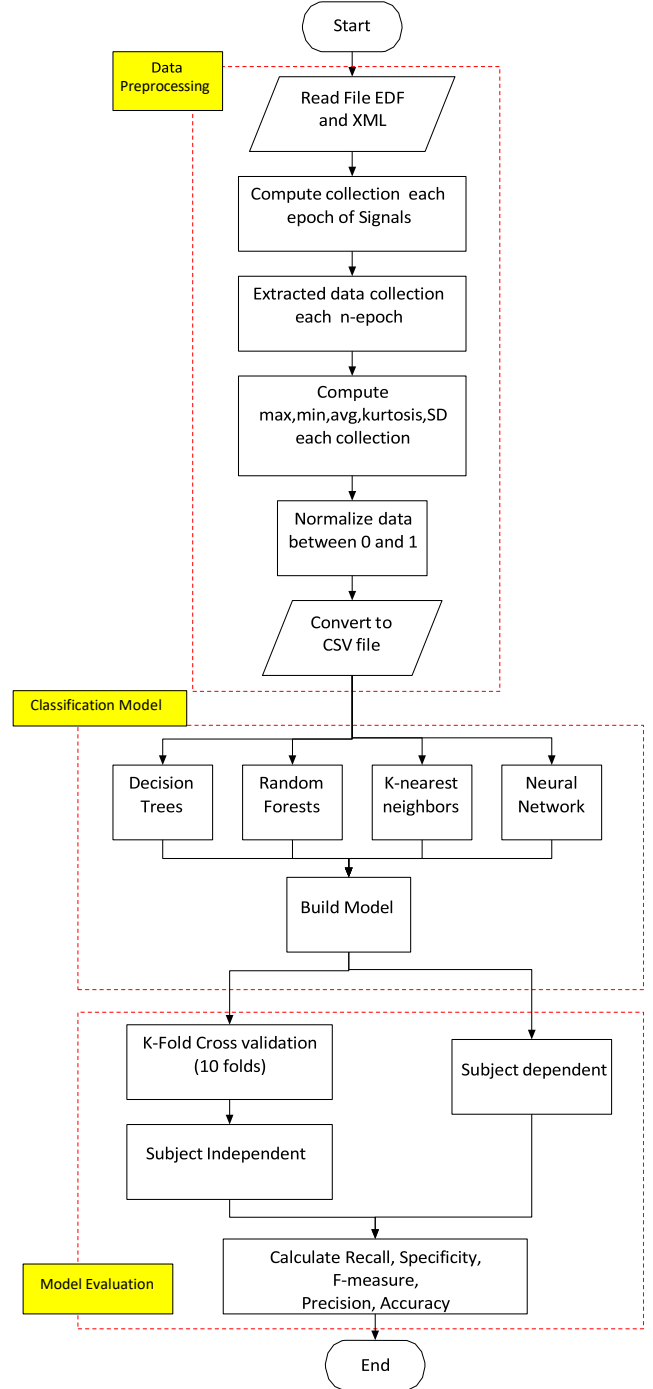


Fig. 5. Overall investigation design.

After extracting data from the dataset as shown in Fig. 3, we selected representative features including maximum, minimum, average, kurtosis, and standard deviation from the extracted signals in different epochs. The followings were the selected feature used in this study.

$$\text{Maximum} = \text{Max}(X_i) \quad (2)$$

$$\text{Minimum} = \text{Min}(X_i) \quad (3)$$

$$\text{Mean}(\bar{x}) = \frac{\sum_{i=1}^N x_i}{N} \quad (4)$$

$$\text{Standard Deviation}(SD) = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}} \quad (5)$$

$$\text{Kurtosis} = \frac{\sum_{i=1}^N \frac{(x_i - \bar{x})^4}{N}}{SD^4} \quad (6)$$

In addition, due to the vast different between the selected features, the statistical normalization was performed to compromise the difference and measure between 0 and 1. After the normalization, it was combined hypnogram data with each epoch from xml annotation file. And it was converted to a CSV file by each epoch as a record until the end of the epoch for each subject.

C. Classification Model and Model Evaluation

In a machine learning workflow, after feature extraction and feature reduction, different classification algorithms are employed to classify the data. In this paper, we have considered four widely used algorithms for this study including Decision Trees (DT), Random Forest (RF), K-Nearest Neighbors (k-NN), and Neural Network (NN). A thorough comparison of these classification models was performed. Technically, the sci-kit-learn library package was employed in Python programming language for the implementation. The model evaluation required a preparation of a training set and testing set by using random sample selection. The k-fold cross-validation that splits our data into different k-subsets (or folds). We used k-1 subsets to train our data and leave the last subset (or the last fold) as a test set. Subsequently, we tested it against the test set. Testing results were done with subject dependent, as shown in Fig. 5 [8].

For the performance evaluation, measurements have been selected each of the techniques according to the sleep stages. The measurements include Accuracy, Precision, Recall, Specificity and F-Measure. The accuracy is a simple ratio of correctly predicted observation to the total observations. Precision shows the ratio of correctly predicted positive observations to the total predicted positive observations. On the other hand, Recall represents the ratio of correctly predicted positive observations to all observations in an actual class. The specificity represents the proportion of actual positive case which are correctly identified. Finally, the F-Measure represents the weighted average of Precision and Recall, which is one of the standard comparative measurements. These selected measurements can be calculated from incremental counts of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) in confusion matrices.

D. Method

In this study, we had EDF and XML from SHHS for 2,535 files. Then used python to extract biological signal data from files by separate signal to epoch 10, 15, 30 seconds of each channel and used statistical for extract features from 14 channel biomedical signals including maximum, minimum, average, kurtosis, and standard deviation according in data preprocessing section. After data preprocessing, we had convert the processed data to csv file. Then split the data as 2 sets are training set for train the data with classification model (DT, RF, KNN, NN) and testing set for evaluation the data after train with classification model. In training set, we use k-fold cross-validation to estimate the classification model on unseen data by selected 10-fold. After train with k-fold we will evaluation of training set and compare evaluation of testing set. In Fig. 2 shown the overall investigation design of this study.

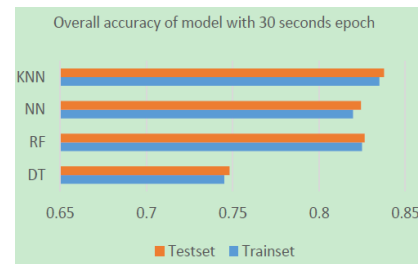


Fig. 6. Overall accuracy of each model with 30 seconds epoch.

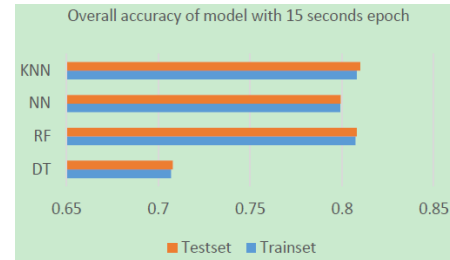


Fig. 7. Overall accuracy of each model with 15 seconds epoch.

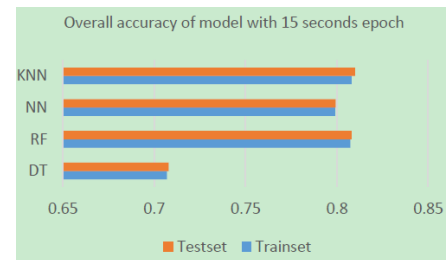


Fig. 8. Overall accuracy of each model with 15 seconds epoch.

TABLE III: OVERALL RESULTS OF 30 SECONDS EPOCH

Measurement/ Model	Decision Trees		Random Forests		Neural Network		K-Nearest Neighbor	
	Trainset	Testset	Trainset	Testset	Trainset	Testset	Trainset	Testset
Accuracy	0.7451	0.7481	0.8248	0.8264	0.8196	0.8242	0.8349	0.8376*
Precision	0.6352	0.6378	0.7391	0.7347	0.7506	0.7569	0.8526	0.8556*
Recall	0.6358	0.6380	0.6683	0.6727	0.6768	0.6829*	0.6737	0.6769
Specificity	0.9288	0.9296	0.9487	0.9496	0.9475	0.9488	0.9506	0.9515*
F-measure	0.6354	0.6378	0.6699	0.6715	0.6983	0.7048*	0.6899	0.6935

TABLE IV: ACCURACY OF EACH SLEEP STAGE CLASSIFICATION IN 30 SECONDS EPOCH

Sleep Stage/ Model	Decision Trees		Random Forests		Neural Network		K-Nearest Neighbor	
	Trainset	Testset	Trainset	Testset	Trainset	Testset	Trainset	Testset
Wake	0.8745	0.8759	0.9192	0.9205	0.9154	0.9172	0.9261	0.9282*
NREM1	0.9421	0.9427	0.9667	0.9671*	0.9642	0.9645	0.9640	0.9646
NREM2	0.8100	0.8120	0.8703	0.8725*	0.8646	0.8663	0.8522	0.8562
NREM3	0.9462	0.9465	0.9659	0.9665*	0.9613	0.9617	0.9496	0.9504
REM	0.9174	0.9191	0.9477	0.9486	0.9442	0.9431	0.9473	0.9490*

TABLE V: OVERALL RESULTS OF 15 SECONDS EPOCH

Measurement/ Model	Decision Trees		Random Forests		Neural Network		K-Nearest Neighbor	
	<i>Trainset</i>	<i>Testset</i>	<i>Trainset</i>	<i>Testset</i>	<i>Trainset</i>	<i>Testset</i>	<i>Trainset</i>	<i>Testset</i>
Accuracy	0.7069	0.7079	0.8073	0.8081	0.7991	0.7993	0.8081	0.8099*
Precision	0.5887	0.5896	0.8146	0.8160*	0.7439	0.7175	0.7164	0.7189
Recall	0.5890	0.5899	0.6334	0.6344	0.6358	0.6363	0.6710	0.6736*
Specificity	0.9181	0.9184	0.9423	0.9426	0.9409	0.9411	0.9450	0.9456*
F-measure	0.5888	0.5897	0.6435	0.6448	0.6378	0.6401	0.6860	0.6887*

TABLE VI: ACCURACY OF EACH SLEEP STAGE CLASSIFICATION IN 15 SECONDS EPOCH

Sleep Stage/ Model	Decision Trees		Random Forests		Neural Network		K-Nearest Neighbor	
	<i>Trainset</i>	<i>Testset</i>	<i>Trainset</i>	<i>Testset</i>	<i>Trainset</i>	<i>Testset</i>	<i>Trainset</i>	<i>Testset</i>
Wake	0.8622	0.8625	0.9104	0.9107	0.9058	0.9078	0.9243	0.9251*
NREM1	0.9384	0.9383	0.9652*	0.9651	0.9648	0.9644	0.9604	0.9603
NREM2	0.7750	0.7762	0.8427	0.8433	0.8388	0.8377	0.8430	0.8445*
NREM3	0.9307	0.9310	0.9563	0.9564*	0.9528	0.9526	0.9437	0.9442
REM	0.9074	0.9078	0.9401	0.9408	0.9360	0.9361	0.9448	0.9455*

TABLE VII: OVERALL RESULTS OF 10 SECONDS EPOCH

Measurement/ Model	Decision Trees		Random Forests		Neural Network		K-Nearest Neighbor	
	<i>Trainset</i>	<i>Testset</i>	<i>Trainset</i>	<i>Testset</i>	<i>Trainset</i>	<i>Testset</i>	<i>Trainset</i>	<i>Testset</i>
Accuracy	0.6931	0.6943	0.7972	0.7980	0.7846	0.7851	0.8061	0.8086*
Precision	0.5733	0.5748	0.8117*	0.8108	0.7593	0.7603	0.7250	0.7278
Recall	0.5742	0.5755	0.6211	0.6222	0.6162	0.6150	0.6615	0.6650*
Specificity	0.9143	0.9147	0.9394	0.9396	0.9364	0.9366	0.9439	0.9447*
F-measure	0.5737	0.5751	0.6327	0.6339	0.6215	0.6213	0.6788	0.6822*

TABLE VIII: ACCURACY OF EACH SLEEP STAGE CLASSIFICATION IN 10 SECONDS EPOCH

Sleep Stage/ Model	Decision Trees		Random Forests		Neural Network		K-Nearest Neighbor	
	<i>Trainset</i>	<i>Testset</i>	<i>Trainset</i>	<i>Testset</i>	<i>Trainset</i>	<i>Testset</i>	<i>Trainset</i>	<i>Testset</i>
Wake	0.8588	0.8597	0.9088	0.9093	0.9016	0.9041	0.9243	0.9253*
NREM1	0.9379	0.9377	0.9652*	0.9651	0.9650	0.9650	0.9629	0.9629
NREM2	0.7619	0.7628	0.8311	0.8320	0.8235	0.8228	0.8380	0.8405*
NREM3	0.9238	0.9240	0.9515	0.9517*	0.9474	0.9476	0.9438	0.9444
REM	0.9038	0.9043	0.9377	0.9379	0.9317	0.9308	0.9432	0.9442*

V. RESULT AND DISCUSSION

This section, the experimental results and the significance of our findings are presented. Fig. 6-8 show all 4 classifiers performance for all sleep stage classification separate by 30, 15 and 10 seconds epoch of each of the selected models. In Table III-VIII shows overall classification results based on the selected measurements in each of the epoch lengths. Furthermore, the classification results on each of the sleep stages are identified. The maximum values in each of the cases are clearly marked. The k-NN achieved the highest correct classification percentage of 83.76%, followed by the RF of 82.64% in the 30 seconds epoch. However, the NN attained the highest F-Measure of 0.7048. In 15 seconds epoch, k-NN also attained the highest accuracy of 80.99%, followed by the RF of 80.81%. However, all measurements show that the k-NN performed promising results. Similar results are found in the 10 seconds epoch. It is seen that KNN and RF perform consistently better in terms of overall accuracy with epoch at 30 seconds. However, upon closer examination of the classification accuracy of individual sleep stages in epoch at 30 seconds, it is seen that k-NN perform very poorly to classify the NREMs sleep stage. On the other hand, NREMs are best classified with RF.

VI. CONCLUSION

In this study, 2,535 subjects were tested using

polysomnographic data with 14 channels of biomedical signals to demonstrate the performance of sleep stage classification using 4 models as Decision Trees, Random Forests, Neural Network, and K-Nearest Neighbors with 3 epochs at 30, 15, 10 seconds. From the experimental results presented in this paper, the epoch at 30 seconds and Random Forests provided the best results compared to other models for classifying sleep stages at 83.76%. It can preliminarily conclude that the epoch at 30 seconds with k-NN is suitable for sleep stage classification purposes.

REFERENCES

- [1] World Health Organization. (February 2018). Noncommunicable diseases country profiles 2018. [Online]. Available: <https://www.who.int/nmh/publications/ncd-profiles-2018/en/>
- [2] D. Brunk, "IOM: It's time to wake up to sleep disorders," *Clinical Psychiatry News*, vol. 34, no. 6, 2006.
- [3] S. Redline, M. Sanders *et al.*, "Methods for obtaining and analyzing unattended polysomnography data for a multicenter study," *Sleep*, Sleep Heart Health Research Group, vol. 21, no. 7, pp. 759-767, 1998.
- [4] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, and C. Shearer. (1999). CRISP-DM 1.0 Step-by-step data mining guide. [Online]. Available: <https://www.the-modeling-agency.com/crisp-dm.pdf>
- [5] S. Khalighi, T. Sousa, G. Pires, and U. Nunes, "Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels," *Xpert Systems with Applications*, vol. 40, no. 17, pp. 7046-7059, 2013.
- [6] D. Zumsteg, H. Hungerbühler, and H. Wieser, *Atlas of Adult Electroencephalography*, Hippocampus Verlag GmbH, 2004.
- [7] O. Kocak, T. Bayrak, A. Erdamar, L. Ozparlak, Z. Telatar, and O. Eroglu, "Automated detection and classification of sleep apnea types using Electrocardiogram (ECG) and Electroencephalogram (EEG)

Features," *Advances in Electrocardiograms - Clinical Applications*, pp. 211-230, 2012.

- [8] J. Han, M. Kamber and P. Jian, *Data Mining Concepts and Techniques*, San Francisco, CA: Morgan Kaufmann Publishers, 2011.
- [9] M. Rahman, M. Bhuiyan and A. Hassan, "Sleep stage classification using single-channel EOG," *Computers in Biology and Medicine*, vol. 102, pp. 211-220, 2018.
- [10] D. Cho and L. Boreom, "Optimized automatic sleep stage classification using the normalized mutual information feature (NMIFS) method," in *Proc. 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Seogwipo, 2017.



Thakerng Wongsirichot is an assistant professor in information technology at Prince of Songkla University, Thailand. He obtained his master of information systems from University of Wollongong, Australia. His main research interests are medical data mining, healthcare information system, health related mobile application, etc.



Nittida Elz is an assistant professor in computer science at Prince of Songkla University, Thailand. She obtained her Ph.D. in computer engineering from La Trobe University, Australia. Her main research interests are network security, network management, etc.



Supasit Kajkamhaeng is a full time lecturer at Prince of Songkla University, Thailand. He obtained his master of engineering in computer engineering from Kasetsart University, Thailand. His main research interests are parallel computing, high performance computing, etc.



Wanchai Nupinit holds a bachelor of science in information and communication technology from Prince of Songkla University, Thailand. He was in the development team of this project. His main interests are data mining and mobile application development.



Narongrit Sridonthong holds a bachelor of science in information and communication technology from Prince of Songkla University, Thailand. He was in the development team of this project. His main interests are data mining and mobile application development.