

Fast Convolutional Method for Automatic Sleep Stage Classification

Intan Nurma Yulita, MT^{1,2}, Mohamad Ivan Fanany, PhD¹, Aniati Murni Arymurthy, PhD¹

¹Machine Learning and Computer Vision (MLCV) Lab, Faculty of Computer Science, Universitas Indonesia, Jawa Barat, Indonesia; ²Department of Computer Science, Universitas Padjadjaran, Sumedang, Indonesia

Objectives: Polysomnography is essential to diagnose sleep disorders. It is used to identify a patient's sleep pattern during sleep. This pattern is obtained by a doctor or health practitioner by using a scoring process, which is time consuming. To overcome this problem, we developed a system that can automatically classify sleep stages. **Methods:** This paper proposes a new method for sleep stage classification, called the fast convolutional method. The proposed method was evaluated against two sleep datasets. The first dataset was obtained from physionet.org, a physiologic signals data centers. Twenty-five patients who had a sleep disorder participated in this data collection. The second dataset was collected in Mitra Keluarga Kemayoran Hospital, Indonesia. Data was recorded from ten healthy respondents. **Results:** The proposed method reached 73.50% and 56.32% of the F-measures for the PhysioNet and Mitra Keluarga Kemayoran Hospital data, respectively. Both values were the highest among all the machine learning methods considered in this study. The proposed method also had an efficient running time. The fast convolutional models of the PhysioNet and Mitra Keluarga Kemayoran Hospital data needed 42.60 and 0.06 seconds, respectively. **Conclusions:** The fast convolutional method worked well on the tested datasets. It achieved a high F-measure result and an efficient running time. Thus, it can be considered a promising tool for sleep stage classification.

Keywords: Machine Learning, Neural Networks, Classification, Sleep Stages, Polysomnography

I. Introduction

Sleep is a vital human need. Through sleep, physical and mental fatigue are relieved. Without adequate sleep, the

ability to concentrate and participate in daily activities is decreased [1]. Research has shown that lack of sleep causes loss of strength, damages the immune system, and increases blood pressure [2,3]. Sleep disorders can be observed through examination of the sleep stage pattern.

Submitted: May 17, 2018

Revised: 1st, July 8, 2018; 2nd, July 22, 2018

Accepted: July 24, 2018

Corresponding Author

Intan Nurma Yulita, MT

Faculty of Computer Science, University of Indonesia, Kota Depok, Jawa Barat 16424, Indonesia. Tel: +62-2178-6341-9, E-mail: intanurma@gmail.com

Polysomnography is a tool to analyze the sleep pattern [4-6]. This test records physical activity when a person asleep. The test is essential as a first step to determine the type of sleep disorder. It combines electromyography (EMG), electrooculography (EOG), electrocardiogram (ECG), electroencephalography (EEG), and so forth. A doctor or health practitioner gives scores based on the data gathered. These scores are the gold standard in sleep stage analysis [7,8].

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

There are several techniques of sleep stage scoring, such as those developed by Rechtschaffen and Kales (R&K) and the American Academy of Sleep Medicine (AASM). The sleep stages are divided into two main categories, namely, rapid eye

movement (REM), and non-rapid eye movement (NREM). In the R&K rules, NREM is divided into four stages: S1, S2, S3, and S4. The S3 and S4 stages can be categorized as slow-wave sleep (SWS). The delta waves comprise less than 50% of the total in S3 and more than 50% in S4. The S4 stage is the deepest among the NREM stages. REM sleep is the deeper than S4 and is often called the dream phase. The AASM rules refer to S1, S2, SWS, and REM as N1, N2, N3, and R.

A doctor or health practitioner takes several days to analyze data gather from a patient. Therefore, there is a need for an application that can automatically classify sleep stages. Several studies have developed machine learning algorithms for sleep stage classification. Aboalayon et al. [9] applied k-nearest neighbors, support vector machines (SVM), and neural networks. The research showed that SVM achieved the best performance. The success factor of SVM was its hyper-planes. Hassan and his colleagues [10,11] performed the ensemble method. Some well-known algorithms for this purpose are boosting and random forest. At the beginning of boosting, all pieces of data have the same weight, but if a particular piece of data has not been classified, its weight is increased in the next iteration. The change in weight causes the data of the minor class to be modeled better. In the random forest method, trees are built, and classification is conducted through the voting of each tree [12]. The ensemble method is a good strategy to handle the class imbalance problem, which often occurs in sleep stage datasets. Langkvist et al. [13] proposed feature learning to achieve highly accurate sleep stage classification. They also implemented hidden Markov models (HMMs) for this purpose.

This paper proposes a new method based on feature learning called the fast convolutional method. It is one-dimensional (1D) convolutional, and this become the basis

of its feature learning. It has an efficient running time and achieved through hierarchical softmax.

II. Methods

Figure 1 illustrates the process of the research design.

1. Signal Data

The dataset was obtained from the PhysioNet data center. The participants included 25 patients (21 men and 4 women) who had a sleep disorder. The patients were aged 28–68 years with body mass index scores ranging from 25.1 to 42.5 kg/m², and their apnea-hypopnea index scores ranged from 1.7 to 90.9. The dataset recorded the movement of the rib cage, oronasal airflow, abdominal movement, snoring, and body position as well as readings obtained by EEG, EMG, EOG, and ECG. However, the PhysioNet data only included EEG, EMG, and EOG readings. The second dataset was obtained from Mitra Keluarga Kemayoran Hospital. Ten healthy respondents participated in this data collection. This study used EEG readings obtained from the Mitra Keluarga Kemayoran Hospital data. Example polysomnography signals are shown in Figure 2.

Every recording was divided into segments by using a 30-second window. An experienced sleep expert reviewed every segment and determined its score. The scoring of the PhysioNet data was based on R&K rules, and the Mitra Keluarga Kemayoran Hospital data was scored according to AASM rules. These scores became the ground truth for our system. Both datasets consisted of 5 labels. The number of segments for each sleep stage are shown in Tables 1 and 2. In the PhysioNet data, SWS and REM were the smallest, which means that the sick patients had difficulty getting deep sleep. In the Mitra Keluarga Kemayoran Hospital data, N1 was the smallest. In normal sleep patterns, N1 is a short stage of just 10 minutes.

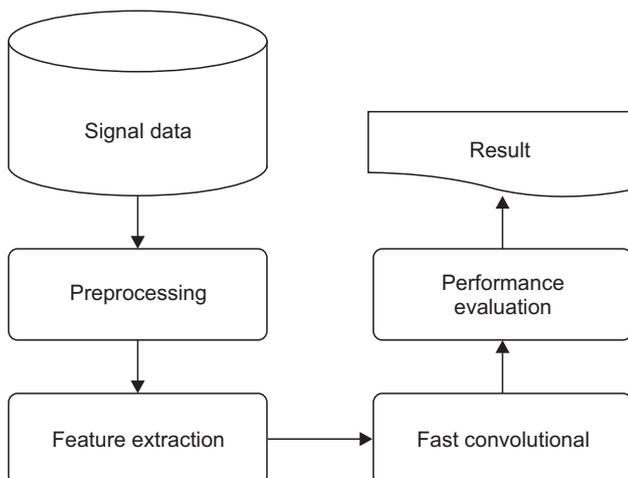


Figure 1. Design of sleep stage classification.

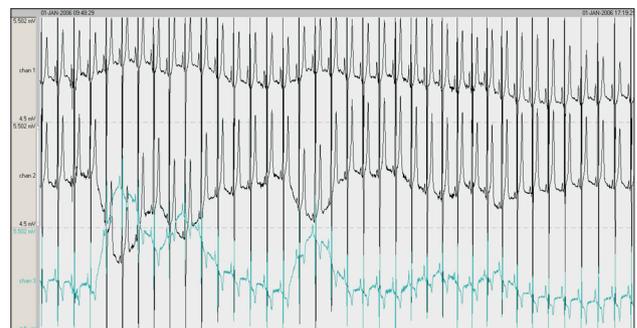


Figure 2. Example polysomnography signals. The first, second, and third signals are electroencephalography (EEG), electrooculography (EOG), and electromyography (EMG), respectively.

Table 1. Number of the segments for each sleep stage and recording on the PhysioNet data

Data	SWS	S2	S1	REM	Awake
1	2,610	5,160	6,390	4,650	3,630
2	4,920	7,620	3,180	5,700	5,040
3	1,560	7,530	2,670	3,750	9,270
4	7,440	2,850	5,430	5,760	2,760
5	3,810	12,420	1,680	3,960	2,520
6	2,490	9,060	2,280	990	7,890
7	4,230	7,320	7,830	2,670	5,700
8	2,370	13,950	3,600	5,070	2,220
9	3,540	8,850	2,670	1,260	10,680
10	4,380	10,110	1,740	5,820	3,870
11	3,330	5,190	4,290	2,010	9,510
12	0	8,010	7,800	2,460	4,950
13	4,380	8,820	6,000	2,040	6,240
14	1,950	11,760	1,080	5,790	3,090
15	4,110	8,400	1,740	570	9,840
16	5,880	10,560	1,410	5,580	2,130
17	1,920	4,650	6,240	4,680	5,070
18	3,810	11,100	3,660	3,840	4,980
19	3,930	6,450	2,100	1,320	9,810
20	2,070	6,810	6,780	1,710	8,460
21	4,110	9,930	3,390	5,040	4,650
22	270	3,720	11,070	1,380	4,890
23	4,170	7,230	3,000	7,380	3,360
24	1,050	15,900	1,860	4,170	3,810
25	1,560	6,150	4,200	2,880	6,840
Total	79,890	209,550	102,090	90,480	141,210

The segments were obtained by dividing each recording by using a 30-second window. Each piece of data consists of segments with their correlated stage labels, namely, slow-wave sleep (SWS), stage 2 (S2), stage 1 (S1), rapid eye movement (REM), and awake.

2. Preprocessing

Preprocessing was conducted to get specific frequencies and amplify other frequencies. This process also reduced the data size without losing information. Notch and bandpass filters were applied to each segment. Notch filtering has the function of removing a single frequency from a signal [14]. This study removed a 50-Hz frequency from the signal at 128 Hz on the PhysioNet data, and at 200 Hz on the Mitra Keluarga Kemayoran Hospital data. The quality factor for the filter was set to 35. A bandpass filter passes a signal in the frequency range above the low-frequency limit (fL) and

Table 2. Number of the segments for each sleep stage and recording on the Mitra Keluarga Kemayoran Hospital data

Data	N1	N2	N3	R	Awake
1	27	307	170	122	234
2	49	396	164	128	114
3	24	244	205	224	114
4	27	389	111	106	138
5	51	249	121	152	208
6	35	271	145	47	186
7	49	359	115	196	65
8	49	413	63	84	119
9	24	253	68	196	114
10	37	408	213	149	94
Total	372	3,289	1,375	1,404	1,386

The segments were obtained by dividing each recording by using a 30-second window. Each piece of data consists of segments with their correlated stage labels, namely, N1, N3, N3, R, and awake.

below the high-frequency limit (fH) [15,16]. This study used $fL = 0.3$ for EEG and EOG, and 10 for EMG. Here, fH was the Nyquist frequency (sample rate/2). This result was then downsampled by two. It caused about 20% of the PhysioNet data and 15% of the Mira Keluarga Kemayoran data, out of the total dataset, to be removed.

3. Feature Extraction

This process was conducted to extract the important feature to distinguish the characteristics of the sleep stages. This step produced 28 features of the PhysioNet data and 17 features of the Mitra Keluarga Kemayoran Hospital data. Tables 3 and 4 show the features.

The frequency domain was required to obtain the EEG signal information. Fast Fourier transform (FFT) was applied to convert the time domain to the frequency domain [17]. FFT utilizes the periodical properties of the Fourier transforms:

$$y_{prel}(fb) = \frac{yp(fb)}{\sum_{f_1}^{f_5} yp(fb)} \quad (1)$$

Here, $y_p(fb)$ is the sum of the absolute power of signal y in frequency band fb . The frequency bands are gamma (20–64 Hz), beta (13–20 Hz), alpha (8–13 Hz), theta (4–8 Hz), and delta (0.5–4 Hz).

The median is the middle value of a signal that has been sorted from the smallest to the largest value. If the amount of

Table 3. Twenty-eight features of the PhysioNet data and their minimum, maximum, mean, and SD values

	Feature	Minimum	Maximum	Mean	SD
1	EEG FFT-delta	-0.05	0.99	0.50	0.19
2	EEG FFT-theta	0.08	1.11	0.49	0.14
3	EEG FFT-alpha	-0.04	1.00	0.50	0.13
4	EEG FFT-beta	-0.11	0.96	0.50	0.13
5	EEG FFT-gamma	-0.03	1.08	0.51	0.13
6	EOG FFT-delta	-0.22	1.06	0.52	0.21
7	EOG FFT-theta	0.20	1.13	0.49	0.13
8	EOG FFT-alpha	0.26	1.39	0.49	0.11
9	EOG FFT-beta	0.24	1.40	0.49	0.10
10	EOG FFT-gamma	0.32	1.42	0.50	0.09
11	EMG FFT-delta	0.45	1.49	0.50	0.06
12	EMG FFT-theta	0.43	1.45	0.50	0.07
13	EMG FFT-alpha	0.24	1.32	0.50	0.09
14	EMG FFT-beta	-0.01	1.11	0.50	0.15
15	EMG FFT-gamma	-0.12	1.00	0.50	0.15
16	EMG median	0.41	1.46	0.50	0.05
17	EOG correlation	-0.33	0.90	0.48	0.27
18	EEG kurtosis	0.15	1.24	0.50	0.09
19	EOG kurtosis	0.07	1.13	0.49	0.11
20	EMG kurtosis	-0.17	1.22	0.50	0.17
21	EOG SD	0.40	1.45	0.50	0.07
22	EEG entropy	-0.43	0.74	0.50	0.09
23	EOG entropy	-0.42	0.82	0.50	0.10
24	EMG entropy	-0.30	1.02	0.50	0.15
25	EEG spectral mean	0.26	1.35	0.51	0.13
26	EOG spectral mean	0.26	1.38	0.49	0.11
27	EMG spectral mean	-0.24	0.91	0.50	0.12
28	EEG fractal exponent	0.04	1.07	0.50	0.13

EEG: electroencephalography, FFT: fast Fourier transform, EOG: electrooculography, EMG: electromyography, SD: standard deviation.

data is odd, then the median value is calculated as

$$Median = \frac{x_{(n+1)}}{2} \tag{2}$$

However, if the amount of data is even, then the median corresponds to Equation (3):

$$Median = \frac{\frac{x_{(n)}}{2} + \frac{x_{(n+1)}}{2}}{2} \tag{3}$$

Table 4. Seventeen features of the Mitra Keluarga Kemayoran Hospital data and their minimum, maximum, mean, and SD values

	Feature	Mini- mum	Maxi- mum	Mean	SD
1	EEG FFT-delta	-0.08	1.11	0.50	0.22
2	EEG FFT-theta	-0.08	1.19	0.50	0.19
3	EEG FFT-alpha	0.16	1.27	0.50	0.18
4	EEG FFT-beta	0.20	1.37	0.50	0.17
5	EEG FFT-gamma	0.21	1.39	0.50	0.20
6	EEG DTCWT level=1 (max)	0.43	1.50	0.50	0.08
7	EEG DTCWT level=1 (min)	-0.50	0.58	0.50	0.08
8	EEG DTCWT level=1 (mean)	-0.43	1.43	0.50	0.04
9	EEG DTCWT level=1 (SD)	0.45	1.49	0.50	0.06
10	EEG DTCWT level=2 (max)	0.44	1.50	0.50	0.08
11	EEG DTCWT level=2 (min)	-0.50	0.57	0.50	0.08
12	EEG DTCWT level=2 (mean)	-0.43	1.43	0.50	0.04
13	EEG DTCWT level=2 (SD)	0.44	1.49	0.50	0.06
14	EEG kurtosis	0.44	1.49	0.50	0.07
15	EEG entropy	-0.37	0.69	0.50	0.13
16	EEG spectral mean	0.07	1.25	0.50	0.21
17	EEG fractal exponent	-0.13	1.15	0.50	0.20

EEG: electroencephalography, FFT: fast Fourier transform, DTCWT: dual-tree complex wavelet transform, SD: standard deviation.

This feature was used to obtain a representation of values that represent two EOG signals (left and right EOG).

Kurtosis is the level of precipitation of a distribution. The feature kurtosis was obtained to recognize the existence of K-complex. K-complex indicates that the data is categorized as S2. Its value is obtained by

$$y_{kurtosis} = \frac{E[y-\mu]^4}{\sigma^4} \tag{4}$$

where μ and σ are the average and standard deviation.

Entropy indicates the randomness of the signal distribution. The more random the distribution, the higher the entropy value. It also indicates signal complexity. The higher the entropy value, the more complex the signal. If a piece of data has a high entropy value, then it has a high probability to be categorized as S1 or REM:

$$y_{entropy} = -\sum_{k=1}^8 \frac{n_k}{N} \ln \frac{n_k}{N} \tag{5}$$

where N is the number of samples in signal y , and n_k is the number of samples of y belonging to the k^{th} bin.

Each signal can be represented by the sum of sinusoids. Its representation is called the signal spectrum. It is often easier to analyze signals regarding spectral representation. The sinusoidal signal represented sleep spindles that indicated N2. The spectral mean value is calculated by

$$Sp_{mean} = \frac{1}{F} \sum_{f1}^{fs} y(fb)_{prel} \cdot fb \tag{6}$$

where F is the sum of the lengths of the 5 frequency bands.

Fractals are a geometric concept in which, if the objects are divided, the parts of the object will be similar to the original object. Its exponent value is calculated as the negative slope in the double logarithmic graph of the spectral density. In this study, if there was no fractal in the signal, then it was categorized as S1 or REM.

Dual-tree complex wavelet transform (DT-CWT) produces the same number of coefficients as the original signal. It is a variation of the discrete wavelet transform (DWT). DT-CWT uses 2 filter trees that represent the real and imaginary parts. The lost signal in the real part can be replaced by the imaginary part. The feature in this study helped to find the K-complex in the signal.

The features of the PhysioNet dataset were normalized by scaling, which made the signal size uniform. Then the under-sampling scheme was applied to handle class imbalance. Some stages had more data than the others. If this problem was not addressed, then the system tended to place each piece of data in the major class. This study implemented under-sampling based on the majority classes. This technique reduced the data in the major classes so that they were

the same as the minority classes. The normalization and under-sampling were conducted in accordance with method the proposed by Langkvist et al. [13] in testing the PhysioNet data.

4. Fast Convolutional Method

Both datasets show that sleep stage patterns appear on some adjacent data. This indicated that the classification model must be capable of capturing correlation among adjacent neighbors. The fast convolutional method does this by using a convolution layer, which is also useful to find the optimal features. The proposed method also had an efficient running time through hierarchical softmax. The softmax architecture is based on the Huffman coding tree [18]. Using this tree, softmax easily finds the class with the highest probability. The convolutional layer and hierarchical softmax are integrated and form the major parts of the proposed method.

Our proposed method is shown in Figure 3. It consists of two primary layers, namely, the feature extraction and classification layers. The feature extraction layer is divided into two sub-layers: convolutional and pooling layers. The fast convolutional method applies a 1D convolutional layer, as shown in Figure 4 [19]. The input of the convolution layer is a 2D feature matrix, $L \times M$. Here, L represents the length of the signal segments and M is the number of features. The window creates a new feature of the data.

At the convolutional layer, each neuron is locally connected to a small input area. All neurons with the same feature map obtain the data from different input areas. Then this layer correlates with some feature maps and also connects to the previous layer. The feature maps are obtained by the kernel. The calculation of the k-feature map (M_{ij}^k) uses the tanh function as follows:

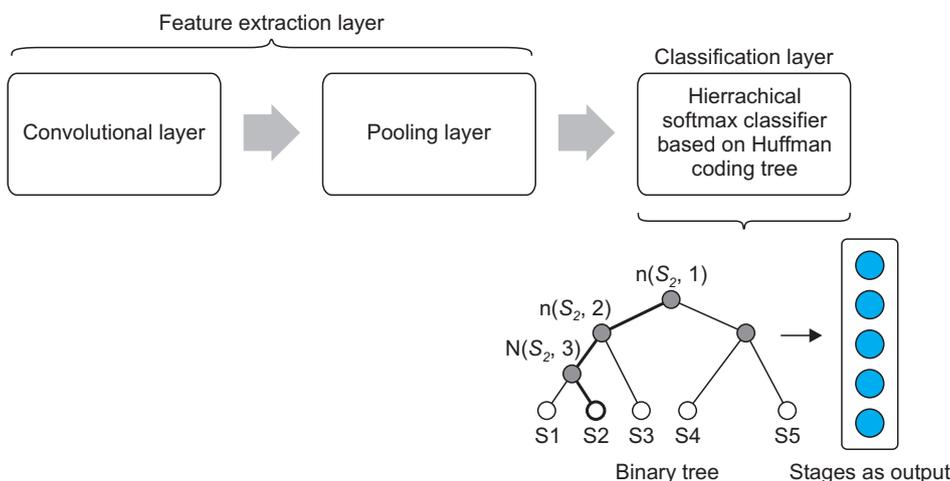


Figure 3. Architecture of the fast convolutional method.

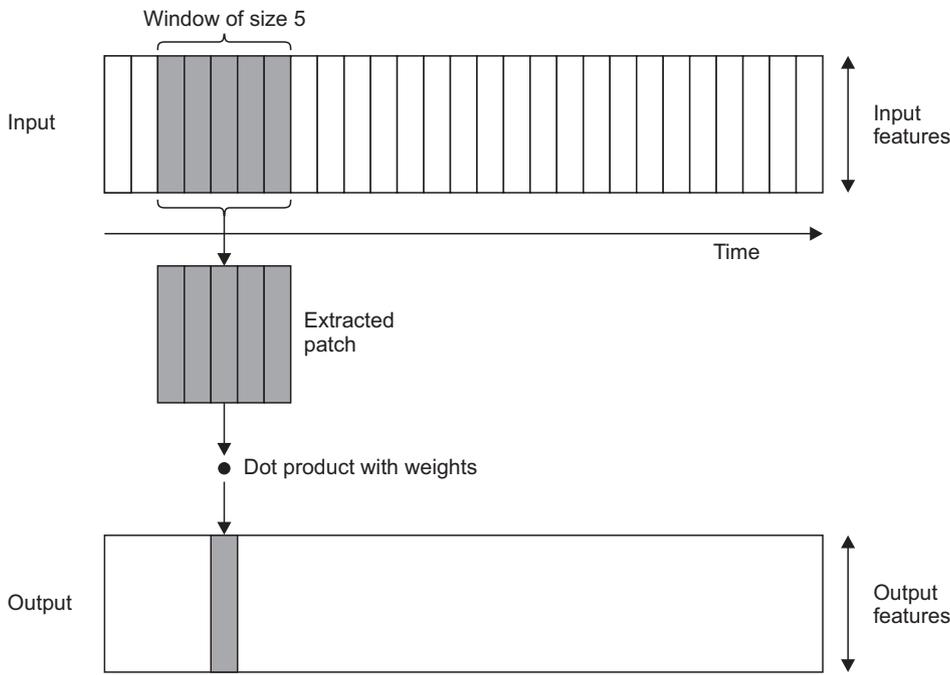


Figure 4. One-dimensional convolution layer.

$$M_{ij}^k = \tanh((W^k \times x)_{ij} + b_k) \tag{7}$$

where W^k = weight, b_k = bias.

The result of this convolutional layer becomes the input for the pooling layer, and it reduces the dimension of feature maps. The common strategies are max pooling and average pooling. The pooling layer divides the output of the convolution layer into several small grids. Then it takes the maximum or average value of each grid, respectively.

Its classification layer implements hierarchical softmax. This softmax enables fast calculation because it depends on fewer parameters. It uses a Huffman binary tree to represent sleep stages and the correlated features, as shown in Figure 5. A leaf represents the sleep stages, and the nodes are the features. For each leaf, there is a unique path from the root to itself, and this path is used to estimate the probability of a sleep stage.

The classification layer has two processes. First, it creates a Huffman tree based on features. Then we calculate the softmax value for each feature and sleep stage as

$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}} \tag{8}$$

Here, f_j is a function of every feature at j for z sleep stage, and k is the number of sleep stages. At the first iteration, its softmax value is calculated for each sleep stage. The feature and its sleep stage with the largest value become a winner

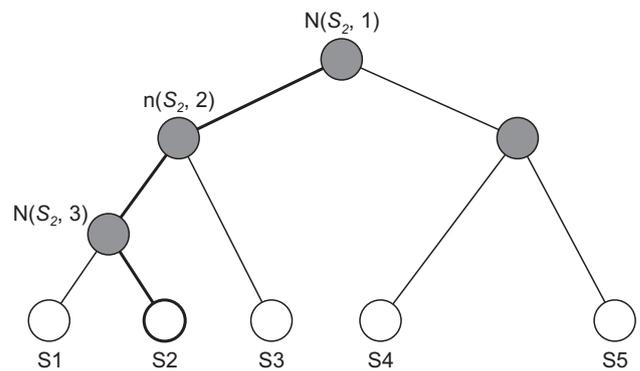


Figure 5. Binary tree of hierarchical softmax.

pair. They have a high priority to be the main node and leaf. In the second iteration, two sub-processes are performed. The first sub-process calculates the softmax value based on the correlation between all features and the winning pair in the previous iteration. The second sub-process calculates the softmax value by computing all features to all sleep stages except the winning sleep stage. In this iteration, there are two winning pairs for each subprocess. The creation of new branches is stopped if all sleep stages have become leaves.

The second process of the classification layer is a prediction process. It is done by calculating softmax values between the data and all sleep stages based on the tree. Figure 5 illustrates the calculation of S_2 . The softmax value is only based on nodes in bold lines. The highest probability of a sleep stage becomes the prediction class. The fast convolutional method computes only the feature of the tree, not all features; thus,

this method has a fast computation.

5. Performance Evaluation

The system was evaluated in terms of precision, recall, F-measure and running time. This study also implemented n -leave-one-out cross-validation, where n represents the number of pieces of data. The proposed method was compared with various machine learning models, such as Bayesian networks, naive Bayes, decision tree, k-nearest neighbors, random forest, bagging, and AdaBoost. We also compared the proposed method and DBN HMM on the PhysioNet data.

III. Results

The activation function in the convolutional layer of the fast convolutional method was based on a tanh function and max-pooling. The windows sizes were 15 in the PhysioNet data, and 3 in the Mitra Keluarga Kemayoran Hospital data. A total of 100 iterations were carried out using root-mean-square propagation as the optimizer. We also compared the proposed method with other machine learning models. The multilayer perceptron used the learning rate = 0.3, momentum = 0.2, and the hidden layers were $(= \text{number of features} + 5)/2$. The decision tree in this study was based on a C4.5 algorithm. The number of neighbors (k) in the k-nearest neighbor method was five. The base classifier of bagging was REPTree (Reduced-Error Pruning C4.5), and AdaBoost was decision stump. They were conducted for 100 iterations.

One important part of SVM is its kernel. This study applied a poly kernel in a multinomial logistic regression model as its calibrator. For comparison, the random forest method was conducted for 100 iterations. The performance of the proposed method was also compared with that of DBN HMM on the PhysioNet data. There were 200 hidden units for the 2 layers of DBN. The DBN had a function of quantization for HMM and was pre-trained for 200 iterations. The top layer used a softmax classifier for 50 iterations. Table 3 shows the results obtained by the proposed fast convolutional method and other methods. Precision, recall, and the F-measure were measured as percentages. The running time was measured in seconds (s). This study was conducted on a PC with Ubuntu 14.04 LTS 64, Hard Disk 240 GB, Four TITAN X GPUs with 12 GB of memory per GPU, Memory DDR2 RAM 64.00 GB, and Core i7-5820K CPU 3.3 @GHz.

IV. Discussion

This chapter discusses the results obtained by each algo-

rithm with the two datasets. Table 5 shows the performance of each algorithm. AdaBoost showed the worst performance. Although it is an ensemble method, it failed to distinguish the stages. On the other hand, bagging and random forest achieved the best performance. The results showed that the decision stump in AdaBoost could not handle the complexity of the dataset. Unlike AdaBoost, the base classifier of random forest was a tree, so its performance was better. The use of resampling in the ensemble method, as well as bagging, also provided good performance. Thus, this algorithm achieved the best performance.

The second best performance was achieved by multi-layer perceptron. The use of multiple layers for feature learning successfully found the optimal feature for classification. This strategy was more effective than the use of hyperplanes in SVM. Naive Bayes had the lowest performance, which indicates that the features were not independent.

The machine learning models had F-measure results below 60% except for bagging and DBN HMM. DBN HMM is a sequence-based classifier. Its classification was not only determined by the features but also the interaction among stages. Both methods had the longest running times, which were more than 3,000 s. This means that it was not easy to get better performance with efficient running time. However, the fast convolutional method was able to do so. The algorithm had the highest F-measure for the PhysioNet data, taking just 42.58 s.

Table 6 presents the confusion matrix of the fast convolutional method for the PhysioNet data. The S2 stage was the most successfully classified. The second and third were S1

Table 5. Performance results on the PhysioNet data

Method	Precision (%)	Recall (%)	F-measure (%)	Training time (s)
Multi-layer perceptron	59.30	59.90	59.00	2,257.80
Decision tree	54.30	54.50	54.40	606.70
K-nearest neighbor	53.40	53.90	53.30	0.10
Bayesian networks	54.00	54.60	53.90	20.90
Bagging	64.20	64.80	64.20	3,586.50
AdaBoost	13.90	35.30	19.90	5.10
Support vector machine	56.90	57.90	57.00	3,229.40
Random forest	64.00	64.60	64.00	617.40
Naive Bayes	52.70	51.30	50.80	4.10
DBN HMM	64.20	62.60	69.10	3,214.00
Fast convolutional	84.00	70.40	73.50	42.60

DBN: dynamic Bayesian network, HMM: hidden Markov model.

and awake, respectively. These stages were the majority classes in the PhysioNet dataset. On the other hand, REM and SWS were minority classes. They were the most challenging labels to be classified. The most difficult class to identify was REM. It was classified as S1 as much as 26.80% and as SWS 15.02%. This indicates that its model was not built well.

The fast convolutional method also performed well with the Mitra Keluarga Kemayoran Hospital data. Table 7 shows that the fast convolutional method had the highest F-measure of 56.32%. Unlike the PhysioNet data, the second highest F-measure was achieved by SVM. The hyper-planes were more needed than the ensemble models. The Mitra Keluarga Kemayoran Hospital data had a clear pattern because the respondents were healthy persons. Thus, hyper-planes could easily separate the stages. SVM needed 8.33 s for training its hyper-planes, which was an efficient running time. However, SVM was outperformed by the fast convolutional method in terms of both F-measure and running time. The fast convo-

Table 6. Confusion matrix of fast convolutional method for the PhysioNet data (unit: %)

	SWS	Stage 2	Stage 1	REM	Awake
SWS	66.90	0.70	3.30	16.10	13.00
Stage 2	0.20	88.70	9.80	0.50	0.90
Stage 1	0.60	7.40	74.40	10.40	7.30
REM	15.00	2.20	26.80	43.10	12.90
Awake	7.90	0.50	11.80	5.50	74.20

Data consist of segments with their correlated stage labels, namely, slow-wave sleep (SWS), stage 2 (S2), stage 1 (S1), rapid eye movement (REM), and awake.

Table 7. Performance results on the Mitra Keluarga Kemayoran Hospital data

Method	Precision (%)	Recall (%)	F-measure (%)	Training time (s)
Multi-layer perceptron	46.36	49.05	47.51	62.39
Decision tree	45.73	43.65	44.63	1.27
K-nearest neighbor	46.63	45.12	45.83	0.01
Bayesian networks	48.41	43.70	45.67	0.33
Bagging	46.99	52.23	49.29	28.95
AdaBoost	47.56	31.92	40.70	40.70
Support vector machine	45.32	65.15	53.41	8.33
Random forest	47.94	46.22	46.96	11.02
Naive Bayes	43.52	34.84	38.42	0.10
Fast convolutional	59.15	61.52	56.32	0.06

lutional models only required 0.06 s to be built.

This study used two different characteristic datasets. This difference was in the segment data distribution for each sleep stage. The data from the healthy participants had a clearer sleep stage pattern than the PhysioNet data. Although the same methods and parameters were used, the models of the machine learning methods for each dataset were different. This means that the chosen machine learning models must suit the characteristics of the dataset they are applied to. However, the proposed method showed good performance in terms of precision, recall, and F-measure for both datasets. It achieved the highest values compared to all machine learning models in this study. Its success can be attributed to its 1D convolutional layer. Its hierarchical softmax method in the classification layer also resulted in efficient running time. It can be concluded that the fast convolutional method is a promising tool for sleep stage classification.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This work is extracted of Intan Nurma Yulita's thesis. It is sponsored by Center of Excellence for Higher Education Research Grant funded by Indonesian Ministry of Research and Higher Education. This paper also backed by GPU Grant from NVIDIA.

References

- Hartzler BM. Fatigue on the flight deck: the consequences of sleep loss and the benefits of napping. *Accid Anal Prev* 2014;62:309-18.
- McEwen BS, Karatsoreos IN. Sleep deprivation and circadian disruption: stress, allostasis, and allostatic load. *Sleep Med Clin* 2015;10(1):1-10.
- Fullagar HH, Skorski S, Duffield R, Hammes D, Coutts AJ, Meyer T. Sleep and athletic performance: the effects of sleep loss on exercise performance, and physiological and cognitive responses to exercise. *Sports Med* 2015;45(2):161-86.
- Weber F, Hoang Do JP, Chung S, Beier KT, Bikov M, Saffari Doost M, et al. Regulation of REM and Non-REM sleep by periaqueductal GABAergic neurons. *Nat Commun* 2018;9(1):354.

5. Begic D. "Polysomnographic" and "sleep" patterns: synonyms or two distinct terms. *Psychiatr Danub* 2015;27(1):73.
6. Kaditis A, Kheirandish-Gozal L, Gozal D. Pediatric OSAS: oximetry can provide answers when polysomnography is not available. *Sleep Med Rev* 2016;27:96-105.
7. de Raaff CA, Pierik AS, Coblijn UK, de Vries N, Bonjer HJ, van Wagenveld BA. Value of routine polysomnography in bariatric surgery. *Surg Endosc* 2017;31(1):245-8.
8. Boostani R, Karimzadeh F, Nami M. A comparative review on sleep stage classification methods in patients and healthy individuals. *Comput Methods Programs Biomed* 2017;140:77-91.
9. Aboalayon KA, Faezipour M, Almuhammadi WS, Moslehpour S. Sleep stage classification using EEG signal analysis: a comprehensive survey and new investigation. *Entropy* 2016;18(9):272.
10. Hassan AR, Bhuiyan MI. Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating. *Biomed Signal Process Control* 2016;24:1-10.
11. Hassan AR, Subasi A. A decision support system for automated identification of sleep stages from single-channel EEG signals. *Knowl Based Syst* 2017;128:115-24.
12. Kuo CY, Yu LC, Chen HC, Chan CL. Comparison of models for the prediction of medical costs of spinal fusion in Taiwan Diagnosis-Related Groups by machine learning algorithms. *Healthc Inform Res* 2018;24(1):29-37.
13. Langkvist M, Karlsson L, Loutfi A. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognit Lett* 2014;42:11-24.
14. Gaur P, Pachori RB, Wang H, Prasad G. A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry. *Expert Syst Appl* 2018;95:201-11.
15. Han D, An S, Shi P. Multi-frequency weak signal detection based on wavelet transform and parameter compensation band-pass multi-stable stochastic resonance. *Mech Syst Signal Process* 2016;70:995-1010.
16. Yener SC, Uygur A, Kuntman HH. Ultra low-voltage ultra low-power memristor based band-pass filter design and its application to EEG signal processing. *Analog Integr Circuits Signal Process* 2016;89(3):719-26.
17. Das A. Fast Fourier transform. In Das A, editors. *Signal conditioning: an introduction to continuous wave communication and signal processing*. Heidelberg: Springer; 2012. p. 193-215.
18. Joulin A, Grave E, Bojanowski P, Douze M, Jegou H, Mikolov T. FastText.zip: compressing text classification models [Internet]. Ithaca (NY): arXiv.org; 2016 [cited at 2018 Jul 15]. Available from: <https://arxiv.org/abs/1612.03651>.
19. Chollet F. *Deep learning with Python*. Shelter Island (NY): Manning Publications Co.; 2018.