Classification of Sleep Stages via Machine Learning Algorithms
Makine Öğrenmesi Algoritmaları ile Uyku Evrelerinin Sınıflandırılması

Ali Bulut[1], Galip Ozturk[1], Ibrahim Kaya[1,*]
[1]Department of Biomedical Engineering, Izmir Katip Celebi University, Izmir, Turkey
ORCIDs: 0000-0002-0325-9195, 0000-0002-4143-8741, 0000-0003-0802-4376
E-mails: aliblt379@gmail.com, galipoozturk@gmail.com, ibrahimkaya21@yahoo.com
*Corresponding author.

***Abstract*—Sleep is a natural form of rest for humans. People need sleep to perform their daily functions. Insufficient or unstable sleep may adversely affect the function of many systems in human body. Sleep disorders can be seen common and cause serious health problems that affect quality of life. From past to present, it has become imperative to classify sleep stages in order to accurately analyze and diagnose these disorders. This classification is made by people who are experts in the field of sleep. However, this process is a very laborious task that requires high attention, and since it is done by a human, it is quite normal to make wrong classifications. As a solution to this, it is possible to make these classifications with machine learning techniques to obtain more accurate results. In this study, we compared different classification methods with each other and examined the channel-based accuracy of the method that gives the highest accuracy based on channels. The accuracy of the Fine Gaussian SVM Method was 98.9% and the F1-score was 98.95, the accuracy of the Weighted KNN Method was 97.9% and the F1-score was 97.89, the accuracy of the Wide Neural Network Method was 97.4% and the F1-score was 97.09, the accuracy of the Cubic SVM Method was 96.2% and the F1-score was 96.36. When we examine the Fine Gaussian SVM Method with the highest accuracy based on channels, we found accuracy of only Fpz-CZ channel is 98.1%, accuracy of only Pz-Oz channel is 94.5%.**

Keywords*—Sleep; sleep stages; machine learning; automatic sleep staging

***Özetçe*—Uyku, insanlar için doğal bir dinlenme halidir. İnsanlar günlük işlevlerini yerine getirebilmek için uykuya ihtiyaç duyarlar. Yetersiz veya dengesiz uyku, insan vücudundaki birçok sistemin işlevini olumsuz etkileyebilir. Uyku bozuklukları yaygın olarak görülebilmekte ve yaşam kalitelerini etkileyen ciddi sağlık sorunlarına neden olmaktadır. Geçmişten günümüze bu bozuklukların doğru bir şekilde analiz ve teşhis edilebilmesi için uyku evrelerinin sınıflandırılması zorunlu hale gelmiştir. Bu sınıflandırma uyku alanında uzman kişiler tarafından yapılmaktadır. Ancak bu işlem oldukça zahmetli ve yüksek dikkat gerektiren bir iştir ve bir insan tarafından yapıldığından yanlış sınıflandırmalar yapılması oldukça normaldir. Buna çözüm olarak daha doğru sonuçlar elde etmek için bu sınıflandırmaları makine öğrenmesi teknikleri ile yapmak mümkündür. Biz bu çalışmada farklı sınıflandırma metodlarını birbiriyle kıyasladık ve en fazla doğruluk veren metodu kanal bazlı inceledik. Fine Gaussian SVM Metodu için doğruluğu 98.9% ve F1-skoru 98.95, Weighted KNN Metodu için doğruluğu 97.9% ve F1- skoru 97.89, Wide Neural Network Methodu için doğruluğu 97.4% ve F1-skoru 97.09, Cubic SVM Metodu için doğruluğu 96.2% ve F1-skoru 96.36 olarak bulduk. En yüksek başarı oranına sahip Fine Gaussian SVM Metodunu kanal bazlı incelediğimizde ise sadece Fpz-CZ kanalının kullanılmasıyla doğruluğu 98.1%, sadece Pz-Oz kanalının kullanılmasıyla doğruluğu 94.5% bulduk.**

Anahtar Kelimeler*—Uyku; uyku evreleri; makine öğrenmesi; otomatik uyku evresi sınıflandırma

I. Introduction

Sleep is a vital process for human health. Adequate and regular sleep is important for a person to lead a better quality of life, both physically and mentally. In cases where sleep is insufficient and irregular, many diseases can occur. Accurate and early diagnosis of these diseases is very important. Sleep staging can be used to diagnose sleep disorders. In order to determine these stages, data are collected from subjects during sleep and these data are evaluated by sleep experts and appropriate classification is made. The most widely used and most effective way to collect these data is polysomnography [1]. Polysomnography (PSG) is the procedure for recording certain data during sleep [2]. With PSG, various data are collected, such as EEG, EOG, EMG, ECG, respiratory conditions, air currents, blood oxygenation, heart rhythm [2]. Among the data collected when sleep stage classification, the most taken into account are the EEG data (Fig. 1).

According to the Rechtschaffen and Kales (R&K) there two main sleep phases NREM and REM, and 6 different stages in total [3]. These stages are called wakefulness (W), non-rapid eye movements (NREM), NREM is divided into 4 different stages in itself (S1-S4), and finally they are called rapid eye movements (REM) (Fig. 2) [4], [5].

On the other hand, according to the latest handbook of the American Academy of Sleep Medicine (AASM) published in 2012, the stages of sleep consist of wakefulness (W), stage I (N1), stage II (N2), stage III (N3) and REM (R). NREM stage

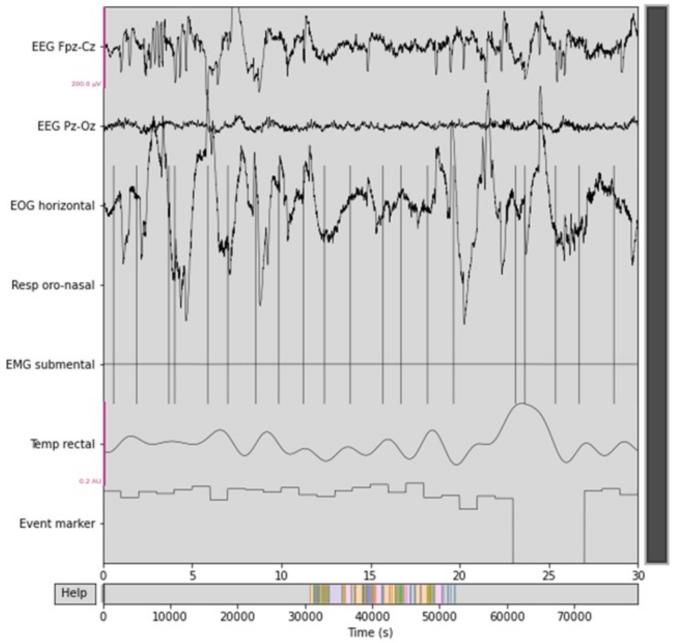

Figure 1: Sample PSG data

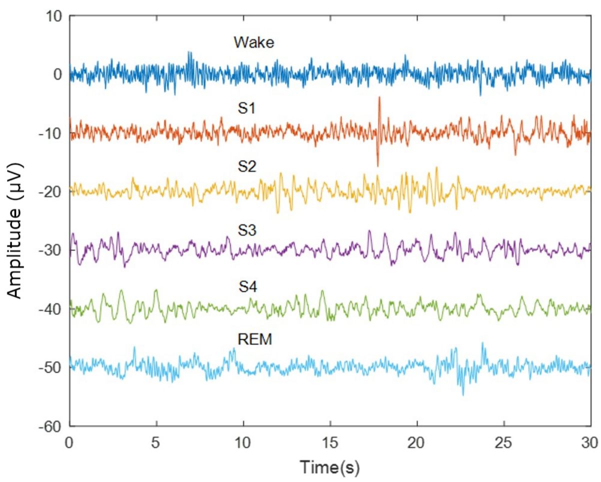

Figure 2: Stages according to the R&K classification [4]

4 has been removed from sleep terminology [3]. According to this manual, it is necessary to pay attention to when scoring sleep stages; 30-second epochs are needed to determine the stages of sleep. Each epoch that exists is named with only one stage. In cases where two stages are located in the same epochs, it is determined which stage is more than half of the epochs, and the epochs are labelled as that stage [3].

When performing sleep staging on a normal PSG recording, various waves and their characteristic properties are considered, as well as many physiological features. These features are; beta activity, theta activity, delta activity, K-complexes, sleep spindles, alpha bursts, amplitude, frequency, EMG level, eye movements [3], [6]. After the EEG data are divided into epochs of the 30s, the stages of sleep are determined by a sleep specialist according to the features they contain in the PSG record. This is a job that requires a very long time and requires a lot of attention, so misclassifications are common [7]. Currently, automatic sleep stage classification systems are being developed to make this staging faster, objective and more accurate. Machine learning-based systems that are trained with attribute extraction along with the above characteristic features give successful results in sleep staging.

II. MATERIALS & METHODS

The data we will use in our study is called sleep-edfx dataset. We got this data from the Physionet database platform, which is a public platform [8], [9]. The sleep-edfx dataset contains 197 whole-night polysomngoraphic recordings, containing EEG, EOG, chin EMG, and event markers. We used just first 20 PSG records of these 197 PSG records in our study. Selected 20 PSG recordings include 10 subjects and each subject has day and night PSG record. The age of the subjects ranged between 25 and 35 years. Sleep stages are scored manually by trained technicians for each PSG recording. Hypnogram files associated with PSG recordings were created. The data epochs are labeled as Wake, REM, Stage 1, Stage 2, Stage 3, Stage 4, M (Movement time), and '?' (not scored) in hypnogram files. The sleep data stages are converted from 6 stage R&K standard to conventional AASM staging standard of 5 sleep stages by combining the stages 3 and stages 4 as N3. In AASM there are Wake, three non-REM N1, N2, N3 stages and a REM stage [4]. In PSG data there are four electrode recordings from two EEG channels Fpz-Cz and Pz-Oz, one EOG and one EMG channels. In addition, there are Resporonasal, EMGSubmenta, Tempbody, and Eventmarker in the data. Two channels of EEG data from Fpz-Cz and Pz-Oz electrodes are used in this study. One sleep EEG epoch contains 3000 time series data points equivalent to 30s at a sampling rate of 100Hz.

The code of the our project developed with Matlab. Only Fpz-Cz and Pz-Oz channel signals are used as the EEG. Firstly program read the single dimensional PSG signals and hypnogram files. After this process, epochs are obtained by sleep states given in hypnogram data. In this codeblock, for machine learning method first the data is split into test and train data [10]. In this method 70 percent of all sleep stages (Wake, N1, N2, N3, REM) for every subjects data is set as train data. This data is selected with permutation for every sleep stage. With this method, the randomization of the training data increased hugely and it create homogeneous randomized training data. After training machine learning algorithm with this data, test part of the process is carried out. Remaining 30 percent data is divided into subject by subject. In this case test data is never used for training. For the machine learning, we use many different features to obtain the highest accuracy as possible. These selected features are given in Table I. After the analysis and emperical method we decided to use; Renyi Entropy, Arithmetical mean, Kurtosis, Logarithmic energy entropy, Variance, Band power alpha, Band power beta, Band power delta, Band power theta, Shannon Entropy and Auto regressive models for Fpz-Cz and Pz-Oz channels. The signal channels (Fpz-Cz and Pz-Oz) are used separately in the codeblock. Every selected features applied separately for Fpz-Cz and Pz-Oz channels so when we use classification learner adds-on in Matlab [11]–[13], we can select the signal channel. It helps us for analyzing channel by channel training and testing and this way we get statistically more information about train and test accuracies.

Feature	Formula
Arithmetic mean	$\overline{x} = \frac{\sum x_i}{n}$
Standard deviation	$S = \sqrt{\frac{\sum \lvert x-\overline{x}\rvert^2}{n}}$
Variance	S^2
Renyi entropy	$\frac{1}{1-\alpha}\ln(\sum_{i=1}^{n} p_i^{\alpha})$
Kurtosis	$\sqrt{\frac{\sum \lvert x-\overline{x}\rvert^4}{nS^4}}$
Logarithmic Energy Entropy	$\log(\sum_{i=1}^{n} s_i^2)$
Band Power Alpha	$\sum_{i=1}^{N} X_{x,alpha}(i)$
Band Power Beta	$\sum_{i=1}^{N} X_{x,beta}(i)$
Band Power Delta	$\sum_{i=1}^{N} X_{x,delta}(i)$
Band Power Tetha	$\sum_{i=1}^{N} X_{x,tetha}(i)$
Shannon Entropy	$-\sum_{i=1}^{N} p_i \log_2 p_i$

Table I: Every features function used into Matlab codeblock [14].

III. RESULTS

The results are evaluated by precision, recall, accuracy, F1-score values that are computed for the test data [15]:

$$Precision = \frac{TP}{FP+TP} \quad (1)$$

$$Recall = \frac{TP}{FN+TP} \quad (2)$$

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} \quad (3)$$

$$F1-Score = 2\cdot\frac{Precision\cdot Recall}{Precision+Recall} \quad (4)$$

where TP = True Positive, FP = False Positive, TN = True Negative, and FN = False Negative.

In addition, we applied the One Way ANOVA Test to the test results obtained for every subject to see the statistical correlations of the channel based models.

The Fine Gaussian SVM model has been the one that has given us the highest accuracy in all the model tests we performed. The accuracy values and F1-scores of the other tested models are indicated in Table II. In order to investigate the effect of channels on accuracy, the Fine Gaussian SVM Method with the largest accuracy value was used. The channel-based accuracy values of the Fine Gaussian SVM Method are as shown in Table III.

Model	Macro F1-Score	Accuracy (%)
Fine Gaussian SVM	0.989	98.9
Weighted KNN	0.979	97.9
Wide Neural Network	0.971	97.4
Cubic SVM	0.964	96.2

Table II: Macro F1- scores and accuracies of the models.

Model	Accuracy (%)
with Fpz-CZ and Pz-Oz	98.9
with Fpz-CZ only	98.1
with Pz-Oz only	94.5

Table III: Channel based accuracies of Fine Gaussian SVM method.

The effect of the channels used on the accuracy of the Fine Gaussian SVM model was studied using test data and the results obtained are as Table **??**.

		Predicted				
		Awake	N1	N2	N3	REM
True	Awake	2141	7	23	8	0
	N1	1	4620	42	11	0
	N2	3	6	6348	27	0
	N3	1	2	61	5867	0
	REM	0	1	25	7	1827

Table IV: The confusion chart of Fine Gaussian SVM method using Fpz-CZ and Pz-Oz channels features altogether.

Sleep Stage	Precision	Recall	F1-Score	Macro Preci-sion	Macro Recall	Macro F1-Score
Awake	0.998	0.983	0.990	0.993	0.987	0.989
N1	0.997	0.988	0.992			
N2	0.977	0.994	0.985			
N3	0.991	0.989	0.989			
REM	1.000	0.982	0.990			

Table V: Classification performances for each sleep stage for Fine Gaussian SVM using Fpz-CZ and Pz-Oz channels features altogether.

It can be clearly seen that when Fpz-Cz and Pz-Oz channels are used, more successful results are obtained. When only the Fpz-CZ channel is used, overall success rate decreased very slightly. But when only the Pz-Oz channel, is used, overall success rate dropped significantly. By looking at these results, it can be inferred that the effect of the Pz-Oz channel on the accuracy value is excessive.

		Predicted				
		Awake	N1	N2	N3	REM
True	Awake	2118	13	25	23	0
	N1	3	4590	52	29	0
	N2	2	21	6309	45	7
	N3	11	46	181	5688	5
	REM	1	10	24	49	1774

Table VI: The confusion chart of Fine Gaussian SVM Method for only Fpz-CZ channel.

Sleep Stage	Precision	Recall	F1-Score	Macro Precision	Macro Recall	Macro F1-Score
Awake	0.996	0.972	0.984	0.986	0.976	0.981
N1	0.986	0.982	0.984			
N2	0.975	0.988	0.981			
N3	0.976	0.985	0.980			
REM	0.996	0.954	0.974			

Table VII: Classification performances for each sleep stage for Fine Gaussian SVM with only Fpz-CZ channel.

Channel-based (Fpz-CZ and Pz-Oz) Fine Gaussian SVM model trained and tested separately on 10 subjects and a total of 20 psg data, day and night. Since the obtained data were normally distributed and the variances of the groups were equal to each other [16]. We applied a one-way ANOVA test to understand whether there is a statistical relationship between the data obtained [13], [17], [18]. As a result of ANOVA, we found the p-value less than 0.001. This result showed that there was a statistically significant relationship between the channel based models. We applied Tukey HSD and Bonferroni tests from Post-hoc tests to see which models this relationship is between. Results are shown in Table X.

		Predicted				
		Awake	N1	N2	N3	REM
True	Awake	1953	47	90	81	8
	N1	10	4398	130	126	10
	N2	4	53	6141	179	7
	N3	11	46	181	5688	5
	REM	1	31	75	68	1685

Table VIII: The confusion chart of Fine Gaussian SVM Method for only Pz-Oz channel.

Sleep Stage	Precision	Recall	F1-Score	Macro Precision	Macro Recall	Macro F1-Score
Awake	0.987	0.896	0.939	0.957	0.933	0.944
N1	0.961	0.941	0.951			
N2	0.928	0.962	0.944			
N3	0.926	0.959	0.942			
REM	0.983	0.906	0.943			

Table IX: Classification performances for each sleep stage for Fine Gaussian SVM with only Pz-Oz channel.

Test	(I) Model	(J) Model	p-value
Tukey HSD	Fpz-CZ and Pz-Oz	Fpz-CZ only	0.612
	Fpz-CZ and Pz-Oz	Pz-OZ only	<0.001
	Pz-Oz only	Fpz-CZ only	0.003
Benferroni	Fpz-CZ and Pz-Oz	Fpz-CZ only	1.000
	Fpz-CZ and Pz-Oz	Pz-OZ only	<0.001
	Pz-Oz only	Fpz-CZ only	0.003

Table X: Tukey HSD and Bonferroni test results for Fine Gaussian SVM classifiers.

IV. Conclusion

This study demonstrates the channel-based effectiveness of the method with the highest accuracy rate on the accuracy of sleep staging, as a result of comparing the success rates obtained from different machine learning classification methods that can be used. In this case, using Fpz-CZ and Pz-Oz channels together for this dataset increases the success rate.

Authors' Contributions

Ali Bulut, preparing and organizing manuscript and performing statistical analysis. Galip Oztürk, feature extraction and selection and codeblock development on matlab. Ibrahim Kaya, finding the dataset and supervision of the research and edits.

References

[1] Spriggs W. Essentials of Polysomnography. Jones & Bartlett Publishers, 2019.

[2] Rundo JV, Downey R. Polysomnography. Book chapter in Handbook of Clinical Neurology, 2019, pp. 160, 381-392.

[3] Kokturk O. Scoring of sleep recordings. Solunum 2013; 15(Suppl. 2): 14-29.

[4] Kaya I. EEG based automatic sleep staging via simple 2D-convolutional neural network. In International Conference on Engineering Technologies (ICENTE'21), November 18-20, 2021, Konya, Turkey.

[5] Hori T, Sugita Y, Koga E, Shirakawa S, Inoue K, Uchida S, Kuwahara H, Kousaka M, Kobayashi T, Tsuji Y, Terashima M, Fukuda K, Fukuda N. Proposed supplements and amendments to 'A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects', the Rechtschaffen & Kales (1968) standard. Psychiatry and Clinical Neurosciences 2001; 55(3): 305–310.

[6] Susmakova K. Human sleep and sleep EEG. Measurement Science Review 2004; 4(2): 59-74.

[7] Isler Y. A Detailed Analysis of the Effects of Various Combinations of Heart Rate Variability Indices in Congestive Heart Failure. PhD thesis at the Department of Electrical and Electronics Engineering, The Graduate School of Natural and Applied Sciences, Dokuz Eylul University, 2009.

[8] Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov PC, Mark R, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000; 101(23): e215-e220.

[9] Kemp B, Zwinderman A, Tuk B, Kamphuisen H, Oberye J. Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG. IEEE Transactions on Biomedical Engineering 2000; 47: 1185–1194.

[10] Isler Y, Narin A, Ozer M. Comparison of the effects of cross-validation methods on determining performances of classifiers used in diagnosing congestive heart failure. Measurement Science Review 2015; 15(4): 196-201.

[11] Degirmenci M, Sayilgan E, Isler Y. Evaluation of Wigner-Ville distribution features to estimate steady-state visual evoked potentials' stimulation frequency. Journal of Intelligent Systems with Applications 2021; 4(2): 133-136.

[12] Altan G, Inat G. EEG based spatial attention shifts detection using time-frequency features on empirical wavelet transform. Journal of Intelligent Systems with Applications 2021; 4(2): 144-149.

[13] Degirmenci M, Yuce YK, Isler Y. Motor imaginary task classification using statistically significant time domain and frequency domain EEG features. Journal of Intelligent Systems with Applications 2022; 5(1): 49-54.

[14] Khalighi S, Sousa T, Pires G, Nunes U. Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels. Expert Systems with Applications 2013; 40(17): 7046-7059.

[15] Zhou D, Wang J, Hu G, Zhang J, Li F, Yan R, Cong F. Singlechannelnet: A model for automatic sleep stage classification with raw single-channel EEG. Biomedical Signal Processing and Control 2022; 75: 103592.

[16] Kul S. Guideline for suitable statistical test selection. Plevra Bulteni 2014; 8(2):26-29.

[17] Akgul A. Tibbi Arastirmalarda Statistiksel Analiz Teknikleri: SPSS Uygulamalari (in Turkish), Seckin Yayincilik, Ankara, Turkey, 2003.

[18] Sayilgan E, Yuce YK, Isler Y. Evaluation of wavelet features selected via statistical evidence from steady-state visually-evoked potentials to predict the stimulating frequency. Journal of the Faculty of Engineering and Architecture of Gazi University 2021; 36(2): 593-605.