DeepUCS for knowledge extraction applied to sleep stages classification*

Rahma Ferjani, Lilia Rejeb, Mohamed Skander TEBOURBI

Université de Tunis, Institut Supérieur de Gestion de Tunis (ISG Tunis)
SMART Lab, Tunis, Tunisia

***Abstract*—For human mental and physical health, sleep is a fundamental restorative process. Sleep analysis is considered as a crucial task to identify the various abnormalities, given the risks that sleep disorders can present. The gold standard for human sleep analysis is sleep scoring. Sleep experts review the PSG recordings and visually identify the various sleep stages for each sleep epoch. Due to the massive volume of recordings acquired during a single sleep period, manual sleep scoring task is considered as a time-consuming and labor intensive task. In this paper we propose a new approach for an interpretable automatic sleep scoring model based on supervised deep learning method and learning classifier system. The effectiveness of our approach was investigated using real electroencephalography (EEG).**

***Index Terms*—Deep learning, Learning classifier system, Convolutional neural network, sUpervised Classifier system, Explainable artificial intelligence.**

I. Introduction

Sleep disorders are one of the most common health issues that are often overlooked. At the same time, they affect health and longevity as being sleep deprived causes cognitive loss. Therefore, they have a negative impact on the essential daily acts such as memory, concentration and moodiness [2]. This is why in research, sleep analysis has become a very important field as well as analyses are considered to be essential tasks for detecting the different anomalies. The best way to analyze human sleep efficiency is by sleep scoring. It consists of identifying the various stages of sleep. Based on polysomnography (PSG) patient recordings obtained at night during sleep.Sleep scoring is carried out manually by experts by reviewing PSG recordings in order to identify sleep stages. This research is always seen as exhausting and time-consuming [1]. Polysomnography is a test to study sleep and find out if or why the patient has experienced sleep disorders. The PSG is a multivariate system consisting of signal recordings such as electroencephalogram (EEG) to monitor brain activities, electrooculography (EOG) to record eye movement, electromyogram (EMG) for muscle activity and electrocardiography (ECG) for heart rhythm monitoring.

II. R&K and AASM sleep standards

The process of sleep scoring is based only on the standard that has been accepted for about approximately 40 years is the Rechtschaffen and Kales (R&K) manual sleep classification (1968). According to R&K standards, sleep is composed in 6 stages (Wakefulness, Stage 1 NREM, Stage 2 NREM, Stage 3 NREM, Stage 4 NREM and rapid eye movement REM). Despite this standard has been useful in many cases, R&K rules have been criticized for their subjective interpretation, which has led to considerable variability in the visual evaluation of sleep stages [7] [9]. In 2007, the American Academy of Sleep Medicine (AASM) proposed a new guide for the classification of sleep stages by amending the standard guide of Rechtschaffen and Kales. According to the AASM, they combined the two phases of sleep S3 and S4 into a single phase called deep sleep and also known as the Slow Waves Sleep (SWS) stage because the characteristics of these 2 stages are very similar [6]. Then the representation is becomes as follows: W (wakefulness), stage N1, stage N2, stage N3, stage REM.

III. Automatic sleep stages classification process

According to the existing literature, the quantitative sleep stage scoring scheme includes 3 common steps, including preprocessing, feature extraction and classification. In addition, in some papers, a feature selection step is added after feature extraction in order to find a suitable feature subset. As shown in 1. In order to strengthen the PSG signals the first step is to eliminate artifacts if present to avoid the misinterpretation of the data and the incorrect result. the second step consists in extracting the appropriate input features from PSG signals.The two primary methods widely studied and used for PSG signal processing are Fast Fourier Transform (FFT) and Wavelet Transform (WT) . Since it is easily applied to non-stationary signals and provides richer information including amplitude, frequency and time, WT is better for PSG processing. The thid step is the classification which consists in classifying the epochs according to their class.

IV. State of the art

Performance indicators, such as classification accuracy, nowadays govern computer-based analysis and classification of physiological signals for applications in health care [15]. The automatic sleep scoring has in recent decades monopolized the interest of authors. Different methods of automatic sleep stage classification usually extract features from the PSG signal to analyze each time period (epoch) and use classification algorithms to determine the sleep stage. For this reason, deep learning represents a major step in understanding physiological signals. Various methods of deep learning have been proposed

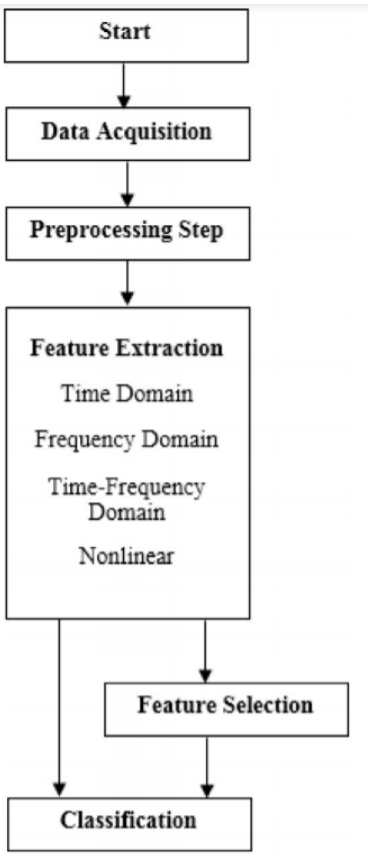

Fig. 1. The process of automatic sleep stages classification [33].

for automatic sleep stages classification and most of them are based on EEG signals since they are classified as the most relevant signals.

Some researchers adopt the use of Convolutional Neural Networks for the automatic sleep stages classification. Yildirim et al. [2] proposed a convolutional neural network method to classify sleep stages and used the EEG signal, EOG signal and 1EEG & 1EOG signals as input to this approach. For the EEG signal, they made a classification of 6 up to 2 classes and they obtained a success rate of 89.43%, 90.48%, 92.24%, 94.23% and 97.85% respectively. In addition they achieved a success rate of 88.28%, 89.77%, 91.88%, 93.76% and 98.06%, respectively for the EOG signal. For a single-channel EEG+EOG they reached 89.54%, 90.98%, 92.33%, 94.34% and 98.06%, respectively. Tsinalis et al. [25] used CNN to learn task-specific filters without using prior domain knowledge based on EEG signal and they obtained 82% as mean accuracy. SORS et al. [26] classified the EEG signal into 5 classes, they developed a network with 14 layers. This approach didn't require a signal preprocessing or feature extraction phase and reached 87% of accuracy. Supratak et al. [27] proposed a DeepSleepNet. The proposed method combined the convolutional neural network and long-short term memory (LSTM) using EEG signal. Accuracy reached 86.2%. Z.Mousavi et al. [13] proposed an approach where the idea is to apply the raw EEG signal directly to the deep convolutional neural network, without involving extraction or selection of features and the classification is made from 2 to 6 classes. The obtained classification accuracies were respectively 98.10%, 96.86%, 93.11%, 92.95%, 93.55%.

Other studies used Recurrent neural Network (RNN) for automatic sleep stages classifications. Hsu et al. [29] proposed an RNN method based on EEG signal energy features to classify different sleep stages. The accuracy rate of this approach is 87.2%. Mousavi et al. [28] developed an approach by combining convolutional neural network with long-short term memory and sequence to sequence (RNN) technique. They used EEG signal as input and they achieved 84.26% of accuracy. In addition, Chen et al. [30] proposed a method based on Hopfield Neural Network (RNN model) and reached an accuracy of 80.6%. Tripathy et al. [31] used an approach founded on auto-encoder for the automated classification of sleep stages. They obtained an average accuracy of 85.51% for 'sleep vs wake' classification, 95.71% for 'light sleep vs deep sleep' and last but not least 95.71% for 'rapid eye movement (REM) vs non-rapid eye movement (NREM)' sleep stages. Based on EOG, Xia et al. [32] proposed a method using Deep Belief Network (RBM model) and achieved 77.7% as an average accuracy.

Though the different techniques mentioned before have provided us with good classification results, they remain uninterpretable due to a peculiar limitation related to the Deep Learning algorithms. This limitation is caused by the Black Box problem. Actually, it is one of the many challenges existing in the Deep learning research activities today as the optimum solution for meeting this challenge is to be able to interpret enigmatic facts about Artificial Intelligence (What we refer to as the Explainable AI). The classification phase was the main focus of our work. Our goal is to make the classification results that are made by the Convulutional Neural Network interpretable through specific rules. To ensure this task, we adopted a combination of CNN and the sUpervised Classifier System. The CNN part of the combination is defined by a neural network model that is mainly used for image classification problems and has shown high accuracy especially in the sleep stages classification problems. The task is therefore performed using polysomnography recordings.

V. Proposed approach DeepUCS

To ensure this task, we adopted a combination of CNN and the sUpervisedClassifier System. The CNN part of the combination is defined by a neural net-work model that is mainly used for image classification problems and has shownhigh accuracy especially in the sleep stages classification problems. This approach contains two parts: A classification and a knowledge extraction. As shown in figure 2, the classification part consists first and foremost in building the model and predicting each 30 seconds image state. Subsequently, we grab the extracted features of each image as well as the result of its related CNN class and use them as an input for the UCS

in order to extract rules which are essential to interpret the classification. As a result, the final outcome is a knowledge base of sleep stages for each 30s epoch and it includes : images, its obtained features and its related extracted rules.

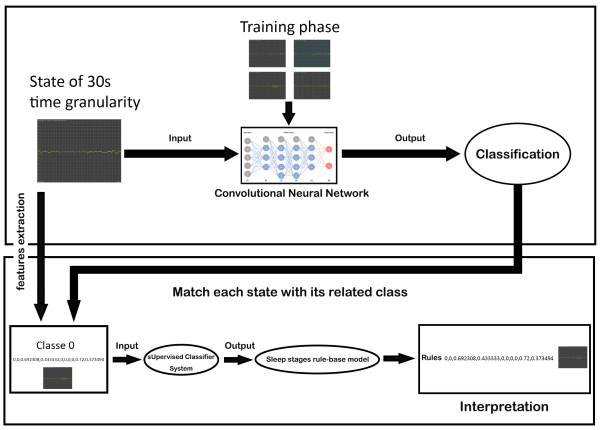

Fig. 2. An overview of knowledge extraction from CNN classification.

VI. Detailed architecture

We start exploring the figure 2 from top to bottom. For the first part of the figure, we trained the model. Then, we predict the class based on the input image. To extract features, each signal is processed with the "Wavelet Transform". A state represents a 30 seconds signal of the image. Therefore, we select the state in addition to its class obtained by the CNN model and use them both as an input for the following second part.

In the second part of the figure 2, we aim to extract rules that allow us to recognize the main criteria for the classification performed by the CNN model, deduct interpretations and select those relevant to our project. Therefore, we use the image state and its corresponding class as an input for the UCS to retrieve knowledge related to the classification results. The final output contains features, rules and the image itself.

1) CNN architecture :: The architecture that we have chosen to apply that is , the CNN algorithm, defined by: (see figure 3)

- 5 layers conv2d : To extract the feature map from the image into a matrix.
- 3 layers maxpooling : That aim to choose the maximum of features so as not to have any loss of information. It serves also to reduce the complexity of the image.
- 1 dropout layer : That eliminates inactive neurons to avoid the overfitting.
- 1 dense layer : That is the final layer and that allows classification.

We used the RELU as the activation function to remove the negative values in the convolutional layer.

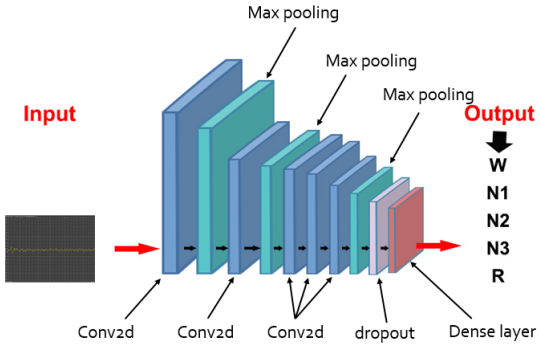

Fig. 3. Our CNN architecture

A. UCS architecture :

Figure 4 explains the flow of the ucs for knowledge extraction. Let's start with the source of data. Each row of the source is made up of a set of features and a corresponding class. The first step is to identifies the rules that respect the conditions of the input to create the matchset without taking into account the Action part. In case where no match between the input and the rule base is found, the covering step comes in to propose rules that correspond to our input. The next step is to select the correct rules in terms of action and put them in the correctset while putting the incorrect rules in the incorrectset. Then, we apply the genetic algorithm on the correct rules after a certain number of iteration that we have it fixed in the parameter to create new adjusted rules using mutation and crossover methods.

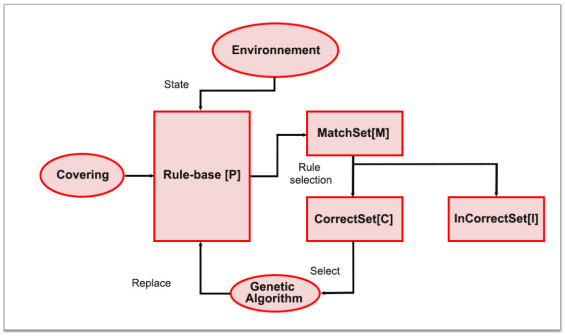

Fig. 4. UCS workflow

VII. Experiments and results

This section presents detailed information on the sleep data sets used in the study and also provides the results of the classification as well as the extracted rules obtained.

A. Dataset description

We used data provided by the PhysioNet site in our experimentation. PhysioNet is an open access platform containing biomedical signals from healthy subjects and patients with different diseases, such as sudden cardiac death, heart

failure, epilepsy, sleep apnea, etc. The data set consists of 153 polysomnographic sleep recordings, including EEG, EOG, chin EMG, some of them also include breathing and body temperature signals. Each recording is accompanied by an annotation file that covers the labels assigned according to the AASM manual by specialists. Sleep stages W, N1, N2, N3 and R, respectively, are represented in the file as 0, 1, 2, 3, 4.

During this project, we used 3 EEG signals as the initial data set. These signals were later on divided into two separate sets: A training set (80%) and a test set (20%).

The descriptions of the dataset is given in the table I where W, N1, N2, N3, R and TNE correspond to the sleep stages classes and the total number of epochs.

	W	N1	N2	N3	R	TNE
Training data	2970	153	948	501	412	4984
Test data	742	38	237	125	103	1245

TABLE I
DESCRIPTION OF THE DATASET

B. Results and discussion

1) Evaluation metrics: In order to evaluate the obtained results, we used some performance metrics: Percentage of Correct Classification (PCC), Confusion matrix, Precision, Recall and F1 measure.

- The PCC is the number of correct classifications divided by the total number of epochs.

$$PCC = \frac{Numberofcorrectclassifications}{Totalnumberofepochs} \quad (1)$$

- The precision$_i$ of a class$_i$ defines how is reliable the model is when it gives us the resulting classification.

$$Precicion_i = \frac{TTP_{all}}{TTP_{all} + TFP_i} \quad (2)$$

With TTP$_all$ is the Total number of True Positive and TFP$_i$ is the Total number False Positive.

$$TTP_{all} = \sum_{j=1}^{n} x_{jj} \quad (3)$$

$$TFP_i = \sum_{j=1}^{n} x_{ji}, with(j \neq i) \quad (4)$$

- The recall$_i$ of a class$_i$ shows how well that class can be detected by the model.

$$Recall_i = \frac{TTP_{all}}{TTP_{all} + TFN_i} \quad (5)$$

With TFN$_i$ is the Total number of False Negative.

$$TFN_i = \sum_{j=1}^{n} x_{ij}, with(j \neq i) \quad (6)$$

- The F1-score$_i$ of a class$_i$ is given by the harmonic mean of the Precision and the Recall which is presented by the equation:

$$F1 - score_i = \frac{2 \times Precision_i \times Recall_i}{Precision_i + Recall_i} \quad (7)$$

As it is a better metric when there are imbalanced classes, as in our case, we decided to focus on the F1-score as an evaluation metric. Imbalanced data distribution exists in most real-life classification problems and thus F1-score is a better metric to test our model.

2) Results and discussion: performance of our CNN model is sensitive to the number of epoch's parameters. We obtained an accuracy of 92.7% after 14 iteration of training.

In table II, we note that the diagonal elements of the confusion matrix represent the number of correct predictions. The first class 'Wake', which stands for the majority class, has 737 correct predictions out of 742 tests done. For the 'N1' class, the result is not accurate. Our data is imbalanced and 'N1' represent a minority class within the data since it contains a small number of epochs. For these mentioned facts, the model does not take into consideration the 'N1'class while learning. We also note that the 'N2' class has 214 out of 237 correct predictions while the 'N3' has 119 out of 125 correct predictions. Finally, the class 'R' was able to correctly predict 99 out of 103.

real classclassified as	W	N1	N2	N3	R
W	737	3	0	0	2
N1	10	6	4	1	17
N2	0	5	214	11	7
N3	1	0	4	119	1
R	0	1	3	0	99

TABLE II
CONFUSION MATRIX OF CNN CLASSIFICATION.

In order to evaluate our classification model, we focus on the F1-score metric. This metric is commonly used for the evaluation of imbalanced datasets, such as in our case.

As seen in the table III, we explored multiple metrics that gave good results related to the existing classes except for the case of the "N1" class. After deciding that our target metric for evaluation is the F1-score for the accurate results that it is known to ensure while dealing with imbalanced datasets, we begin now to inspect the given results. The class "W" has 98.91%, meaning that the model in unlikely to misclassify this class and has properly learned from it instead. We also note that the results are relatively satisfying for the classes "N2", "N3" and "R" with respectively, 92.63%, 92.96% and 86.45%. Now to what concerns the class "N1", the results are unsatisfying. The F1-score corresponding to this class is 22,63% which is way lower than the other mentioned results. We can explain this by the fact that we had few images associated with this class in the training phase.

To extract knowledge from the CNN classification, as we have already mentioned we will based on the UCS, After effectuating several experiments, we fixed our parameters as following:

- Population size: 3500
- Exploration/exploitation rate: 0.5

ClassMetrics	PCC	Precision	Recall	F1-score
W	99.32%	98.52%	99.32%	98.91%
N1	15.7%	40%	15.78%	22.63%
N2	90.29%	95.11%	90.29%	92.63%
N3	95.2%	90.83%	95.2%	92.96%
R	96.11%	78.57%	96.11%	86.45%

TABLE III
ASSESSMENT OF EACH CLASS.

- Number of iterations: 100000
- Mutation probability: μ = 0.06
- Crossover probability: χ = 0.6
- Genetic algorithm: θ_{GA} = 50

Figure 5 is an example of a rule deducted by the UCS for the class W. Starting from this feature [0,0,0.692308,0.433333,0,0,0,0,0.72,0.373494], we were able to extract the following knowledge: When a new state is received and each value in the feature belongs to its corresponding interval, then the probability of being associated with the class W is equal to the accuracy set by the rule.

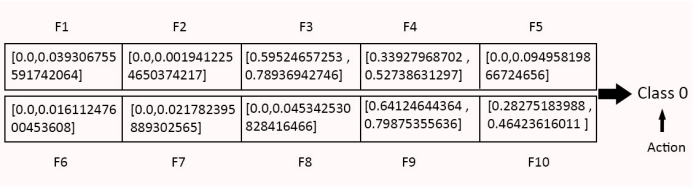

Fig. 5. Example of rule for the class W

.

VIII. STATISTICAL ANALYSIS ON UCS RESULTS

During the knowledge extraction phase, we were able to extract 3294 rules out of 1245 CNN classifications. We decided that rules reaching 50% threshold are considered as reliable rules. Table IV illustrates the number of the reliable rules that we have selected.

ClassRules	reliable rules
W	541
N1	3
N2	78
N3	43
R	24

TABLE IV
STATISTICAL RESULTS

The table IV shows us the distribution of the rules. As it is demonstrated, the majority of the reliable rules belong to the W class. This result was already expected since these rules represent the majority of the UCS states with a 78.52% (541 out of 689 rules). On the other side, the number of rules representing the classes N1, N2, N3 and R are respectively 3, 78, 43, and 24. We are going to explore the result of the R class. Even though the CNN classification for this class was satisfying, we were not able to extract many reliable rules. This fact is due to the inaccurate result of the CNN classification for the class N1. The result related to N1 reduced the performance of class R in the knowledge extraction part by giving it false inputs. This means that the UCS took a lot of features from N1 with class R as input.

IX. CONCLUSIONS

In this paper, we focused on the interpretability of our Deep Learning model and for that specific reason we proposed DeepUCS combination by using Convolutional Neural Networks as a deep learning method. Our target was to interpret the classification results obtained by the CNN based on a set of rules. Since brain activity presents the most important bio-physiological variations that are relevant to the sleep analysis, we used images from real electroencephalography signals for the classification phase as an input. On the other hand, knowledge extraction required features based on the existing signals and the CNN classification as an input for the UCS. Its main goal was to retrieve knowledge allowing us to get the class from a new feature and adding an explanation on how it was classified in the first place. Getting this knowledge was done using our pre-defined rules that have been extracted in a previous step. Finally, we obtained the knowledge related to the sleep stages classification which was inaccessible before due to the Black Box problem caused by the deep learning method that we have opted for. This project aimed to extract knowledge by combining two techniques and it was an attempt to apply the explainable Artificial Intelligence to ensure interpretability of the model by knowledge retrieved. This knowledge and interpretations were necessary in order to reveal what will be later essential to improve the sleep analysis which will result in suggesting the appropriate treatments. As a future work, we aim to use a larger data set to solve the imbalanced characteristic of the classes obtained. It is very crucial point since data set having skewed class proportions affect our model's performance. Therefore, the training data have to represent all classes equally to retrieve more knowledge and avoid struggling to class the new observations. In another perspective, we plan to consider other PSG signals for sleep scoring in addition to the electroencephalography signals that were used in this project. Other PSG physiological signals are available such as electromyography, electrooculography and electrocardiography.

REFERENCES

[1] Pandi-Perumal, S. R., BaHammam, A. S., Brown, G. M., Spence, D. W., Bharti, V. K., Kaur, C., Hardeland, R., & Cardinali, D. P. (2013). Melatonin antioxidative defense: therapeutical implications for aging and neurodegenerative processes. *Neurotoxicity Research*, 23(3), 267–300.

[2] Yildirim, O., Baloglu, U. B., & Acharya, U. R. (2019). A deep learning model for automated sleep stages classification using PSG signals. *International Journal of Environmental Research and Public Health*, 16(4), 599.

[3] O'Reilly, C., Gosselin, N., Carrier, J., & Nielsen, T. (2014). Montreal Archive of Sleep Studies: an open-access resource for instrument benchmarking and exploratory research. *Journal of Sleep Research*, 23(6), 628–635.

[4] Tzimourta, K. D., Tsilimbaris, A., Tzioukalia, K., Tzallas, A. T., Tsipouras, M. G., Astrakas, L. G., & Giannakeas, N. (2018). EEG-based automatic sleep stage classification. *Biomedical Journal*, 1, 6.

[5] Penzel, T., Kantelhardt, J. W., Lo, C.-C., Voigt, K., & Vogelmeier, C. (2003). Dynamics of heart rate and sleep stages in normals and patients with sleep apnea. *Neuropsychopharmacology*, 28(1), S48–S53.

[6] Qureshi, S., Karrila, S., & Vanichayobon, S. (2018). Human sleep scoring based on K-Nearest Neighbors. *Turkish Journal of Electrical Engineering & Computer Sciences*, 26(6), 2802–2818.

[7] Novelli, L., Ferri, R., & Bruni, O. (2010). Sleep classification according to AASM and Rechtschaffen and Kales: effects on sleep scoring parameters of children and adolescents. *Journal of Sleep Research*, 19(1p2), 238–247.

[8] Cao, W., Wang, X., Ming, Z., & Gao, J. (2018). A review on neural networks with random weights. *Neurocomputing*, 275, 278–287.

[9] Moser, D., Anderer, P., Gruber, G., Parapatics, S., Loretz, E., Boeck, M., Kloesch, G., Heller, E., Schmidt, A., Danker-Hopfe, H., et al. (2009). Sleep classification according to AASM and Rechtschaffen Kales: effects on sleep scoring parameters. *Sleep*, 32(2), 139–149.

[10] Marco, L., & Farinella, G. M. (2018). *Computer Vision for Assistive Healthcare*. Academic Press.

[11] Diez, P. (2018). *Smart Wheelchairs and Brain-computer Interfaces: Mobile Assistive Technologies*. Academic Press.

[12] Amin, H. U., Mumtaz, W., Subhani, A. R., Saad, M. N. M., & Malik, A. S. (2017). Classification of EEG signals based on pattern recognition approach. *Frontiers in Computational Neuroscience*, 11, 103.

[13] Mousavi, Z., Rezaii, T. Y., Sheykhivand, S., Farzamnia, A., & Razavi, S. N. (2019). Deep convolutional neural network for classification of sleep stages from single-channel EEG signals. *Journal of Neuroscience Methods*, 324, 108312.

[14] Krogh, A. (2008). What are artificial neural networks? *Nature Biotechnology*, 26(2), 195–197.

[15] Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine*, 161, 1–13.

[16] Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, 187, 27–48.

[17] Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G.-Z. (2016). Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1), 4–21.

[18] Carrio, A., Sampedro, C., Rodriguez-Ramos, A., & Campoy, P. (2017). A review of deep learning methods and applications for unmanned aerial vehicles. *Journal of Sensors*, 2017.

[19] Venkatachalam, M. (2019). Recurrent Neural Networks. Retrieved from https://towardsdatascience.com/recurrent-neural-networks-d4642c9bc7ce accessed 21 March 2020.

[20] Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*.

[21] Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14-15), 2627–2636.

[22] Fischer, A., & Igel, C. (2014). Training restricted Boltzmann machines: An introduction. *Pattern Recognition*, 47(1), 25–39.

[23] Le Ly, D., & Chow, P. (2010). High-performance reconfigurable hardware architecture for restricted Boltzmann machines. *IEEE Transactions on Neural Networks*, 21(11), 1780–1792.

[24] Karhunen, J., Raiko, T., & Cho, K. H. (2015). Unsupervised deep learning: A short review. In *Advances in Independent Component Analysis and Learning Machines* (pp. 125–142). Elsevier.

[25] Tsinalis, O., Matthews, P. M., Guo, Y., & Zafeiriou, S. (2016). Automatic sleep stage scoring with single-channel EEG using convolutional neural networks. *arXiv preprint arXiv:1610.01683*.

[26] A. Sors, S. Bonnet, S. Mirek, L. Vercueil, and J.-F. Payen, *A convolutional neural network for sleep stage scoring from raw single-channel EEG*, Biomedical Signal Processing and Control, vol. 42, pp. 107–114, 2018.

[27] A. Supratak, H. Dong, C. Wu, and Y. Guo, *DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG*, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 11, pp. 1998–2008, 2017.

[28] S. Mousavi, F. Afghah, and U. R. Acharya, *SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach*, PloS one, vol. 14, no. 5, 2019.

[29] Y.-L. Hsu, Y.-T. Yang, J.-S. Wang, and C.-Y. Hsu, *Automatic sleep stage recurrent neural classifier using energy features of EEG signals*, Neurocomputing, vol. 104, pp. 105–114, 2013.

[30] X. Chen, B. Wang, and X. Wang, *Automatic sleep stage classification for daytime nap based on hopfield neural network*, in *Proc. 25th Chinese Control and Decision Conference (CCDC)*, pp. 2671–2674, 2013.

[31] R. K. Tripathy and U. R. Acharya, *Use of features from RR-time series and EEG signals for automated classification of sleep stages in deep neural network framework*, Biocybernetics and Biomedical Engineering, vol. 38, no. 4, pp. 890–902, 2018.

[32] B. Xia, Q. Li, J. Jia, J. Wang, U. Chaudhary, A. Ramos-Murguialday, and N. Birbaumer, *Electrooculogram based sleep stage classification using deep belief network*, in *Proc. International Joint Conference on Neural Networks (IJCNN)*, pp. 1–5, 2015.

[33] R. Boostani, F. Karimzadeh, and M. Nami, *A comparative review on sleep stage classification methods in patients and healthy individuals*, Computer Methods and Programs in Biomedicine, vol. 140, pp. 77–91, 2017.

[34] P. Huilgol, *Accuracy vs. F1-Score*, available at: https://medium.com/analytics-vidhya/accuracy-vs-f1-score-6258237beca2, accessed December 14, 2020.