Automatic Detection of Sleep Stages Using Deep Learning Algorithm

Prakruthi Manjunatha
Dept. of Medical Electronics Engg.,
Ramaiah Institute of Technology
Bangalore, India
prakruthim21@gmail.com

Vibha Chaithanya Rangappa
Dept. of Medical Electronics Engg.,
Ramaiah Institute of Technology
Bangalore, India
vibha.rangappa@gmail.com

Abhishek Varati
Dept. of Medical Electronics Engg.,
Ramaiah Institute of Technology
Bangalore, India
1ms19ml002@gmail.com

C K Narayanappa
Dept. of Medical Electronics Engg.,
Ramaiah Institute of Technology Bangalore, India
c_k_narayanappa@msrit.edu

2022 4th International Conference on Circuits, Control, Communication and Computing (I4C) | 979-8-3503-9747-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/I4C57141.2022.10057685

***Abstract*—Sleep plays a vital role in an individual's life, it enables proper cognitive and behavioral functions. An insufficient amount of sleep can lead to serious repercussions. Sleep disorders adversely affect a person's lifestyle and poses added risk of disorders such as diabetes, high blood pressure, heart disease and stroke, mood disorders, weight gain and obesity. Manual sleep staging and analysis is burdensome and time consuming, hence automatic sleep staging is required. Sleep is classified into various stages (Awake, N1, N2, N3 and REM). The aim of this paper is to survey the progress and challenges in various existing Electroencephalogram (EEG) signal-based methods used for sleep staging in an attempt to find the research gaps and possibly introduce a reasonable solution. Thus, in this paper, a novel and efficient technique to identify sleep stages using Convolutional Neural Network(CNN), CNN + LSTM (Long Short Term Memory) methods applied to 30s epochs of single-channel Fpz-Cz channels. In this study, the PhysioNet sleep cassette study data is used in European Data Format (EDF) Database was used. The proposed methodology achieves accuracy:84.19%, MF1:76.76% respectively.**

Keywords—Sleep analysis, CNN, LSTM, BiLSTM, N1, N2, REM, hypnogram

I. INTRODUCTION

The immense existing evidence concerning the health benefits of sleep on health, cognitive ability and emotional stability, emphasis on sleep study and its benefits has significantly increased [1]. There have been various studies conducted that have concluded in a significant decrease in sleep quality among students and working professionals that could ultimately lead to decline in health and work productivity [2]. However, it is costly and inefficient to manually classify sleep stages, thus the need for an automatic classifier arises.

Sleep is split into five stages according American Academy of Sleep Medicine (AASM), these stages are Wake (W), Non-REM1 (N1), Non-REM 2 (N2), Non-REM 3 (N3) and REM (R) [3]. During W-stage, alpha activity (8-13 Hz) becomes prominent. N1 is a transitional stage between W and N2. This stage is characterized by the appearance of theta activity (4-8 Hz) and the loss of alpha activity [4]. N2 is the stage where the actual sleep begins, this stage produces a unique frequency waveform called the sleep spindle (12-15 Hz) [5]. N3 is considered as deep sleep because the function of the brain is significantly reduced. In this stage, mainly the delta wave (0.5-4 Hz) with the strongest amplitude appears [6].

The present paper explains a method of classification based on CNN and LSTM that uses only the EEG Fpz-Cz channel. The network's architecture, based on the approach from [9], plus a novel set of hyperparameters and training technique allow to achieve results (ACC: 84.19%, MF1: 76.76%) comparable with the current state of the art. This study would help identify optimal features associated with sleep for automatic sleep stage classification.

II. MATERIALS AND METHODS

The following section examines the theoretical perspective on sleep. as well as its stages, the dataset to be utilized in the experiments, design of the two classifiers to be tested and their experimental design for training and testing them. Finally, the technical specifications for implementation are outlined.

A. Dataset

The dataset used for this study is extracted from the Sleep-EDF's "Sleep Cassette Study Data" which is a result of the study conducted by Kemp et al [7]. Albeit the recordings include various signals from the subjects only the EEG Fpz-Cz is considered. Different studies have proved that using only this channel can be enough to classify sleep with human-level accuracy [9]. The dataset includes 153 sets of PSG and hypnogram recordings where 145 of the same was considered as training and the rest was test data.

B. CNN Model

The first experiment in the study employs a CNN to extract features from a concatenation of neighbouring epochs (30 second sleep periods) and a classification of neighbouring epochs (30 second) and a classification with a final Softmax layer. This approach is based on the findings of Chambon et al[8], who demonstrated that when a sleep classifier has access to features from previous and future epochs for a given epoch, its performance improves. The CNN architecture is depicted in Figure 1. There are two main branches defined, one branch learns temporal information from the waves represented in the time series using short filters and strides, while the other learns frequency information using longer filters and strides. Each branch, on its own, has the next sequence:

1DCONV→MP→DROP→1DCONV→1DCONV→1DCONV →MP.

Each 1D convolution layer has 3 steps: convolution, batch normalisation, and a ReLU activation function. Ultimately, the outputs from both branches are flattened and concatenated before being fed into a Softmax layer of 5 units (one for each possible sleep stage).

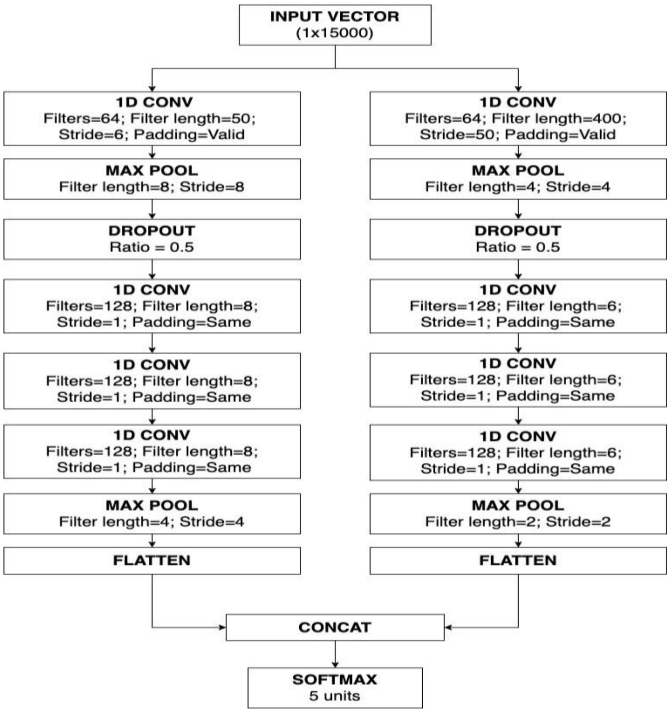

Figure 1. CNN model architecture

C. CNN + LSTM Model

The second experiment involved the use of a combination of CNN and LSTM. This experiment's results are compared with the first experiment's results to determine which of the two strategies, CNN with vector concatenation or CNN+LSTM, is more efficient at learning the temporal dependencies for sleep stage classification.

The same CNN architecture shown in Figure 2 is used for this experiment, but the inputs and outputs have been altered. Each data point is directly fed into the CNN in this case, with no prior concatenation with adjoining epochs. In addition, rather than using the Softmax layer outputs, the activations from the penultimate layer are stored in lists and used as features vectors from each datapoint. As part of this experiment, because the CNN is primarily used as a feature extractor, the Softmax layer activations are unnecessary. The previous layer's activations, on the contrary, serves as a summary of the information extracted from each raw data point. As a result, these activations are regarded as the feature vectors of each data point and are then used as input vectors for the LSTM. The shape of the feature vectors is of 1 × 2688. Since each data point is fed into the CNN independently, the resulting feature vectors are then arranged sequentially in 5 by 5 timesteps to obtain new input vectors of shape 5x2688 for the LSTM component. The LSTM architecture is composed of a main sequence of layers:

DROP→BiLSTM→DROP→BiLSTM→DROP.

A fully connected (FC) layer processes the output of the first dropout layer in parallel. Then, to connect the two paths, a shortcut connection adds the outputs of the first dropout layer and the FC layer. The addition's result is then followed by the final sequence of DROP→FLATTEN and Softmax as a final layer for classification.

Figure 3 depicts the CNN + LSTM ensemble architecture with detailed parameters. Bidirectional LSTM cells are utilized because the network must be able to learn from temporal dependencies in both directions because sleep recording classification is accomplished at the end of the night, when the EEG signal for the complete sleep duration is available. Moreover, each of the system's components is trained independently, i.e., the CNN is used as a feature extractor and the LSTM is used as a classifier.

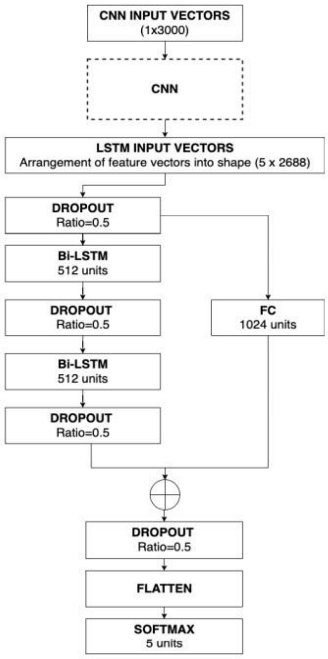

Figure 2. CNN+LSTM model architecture

D. Experimental design and Metrics

Both experiments have used a 20-fold stratified cross validation method based on the Scikit-learn implementation for training and validation. In each fold, the models are trained on 95% of the dataset and then used to make predictions on the remaining 5%. Following the 20 iterations, all fold predictions are concatenated into a Y vector and compared to the Y vector of true values to estimate performance metrics.

The main metrics used to compare the classifiers are: accuracy (ACC), precision per class (Pc), recall per class (Rc), F1 score per class (F1c) and Macro-F1 score (MF1) are mentioned below. Time is also measured in minutes and seconds (mm:ss) to compare training and classification speed.

$$F1_c = \frac{2}{\frac{1}{P_c} + \frac{1}{R_c}} \quad (1)$$

$$ACC = \frac{\sum_{C=1}^{C} TP_c}{N} \quad (2)$$

$$MF_1 = \frac{\sum_{C=1}^{C} F1_c}{C} \quad (3)$$

$$P_c = \frac{TP_c}{TP_c + FP_c} \quad (4)$$

$$R_c = \frac{TP_c}{TP_c + FN_c} \quad (5)$$

where TPc, FPc, FNc are the true positives, false positives and false negatives of class C, respectively; and N is the number of all the data points in the dataset. Where F1c is the F1 score of class c and C is the number of classes, in this case sleep stages.

III. RESULTS

This section describes the most significant findings. These include the data processing results and the two proposed classifiers. The classifiers' performance metrics are also compared to the present state of the art and human-level performance.

A. CNN Model

Minimizing the categorical cross entropy was used to train the CNN with concatenated epochs (CNN Concat). The Adam Optimizer was used with the following parameters: lr = 0.001, 1 = 0.9, 2 = 0.99. The training was done in small batches of 100 data points each, distributed out over 6 epochs. The CNN classifier's confusion matrix is depicted in Figure 4. The classifier has an 81.78% accuracy and a 75.79% MF1 score. The F1 score per class of the wake stage is 88.93%, N1 stage is 40.93%, N2 stage is 84.91%, N3 stage is 87.62%, N4 stage is 74.4%. The average training time for each fold was 0:58 (0 minute and 58 seconds), with a prediction time of less than 1 second for a new sleep recording.

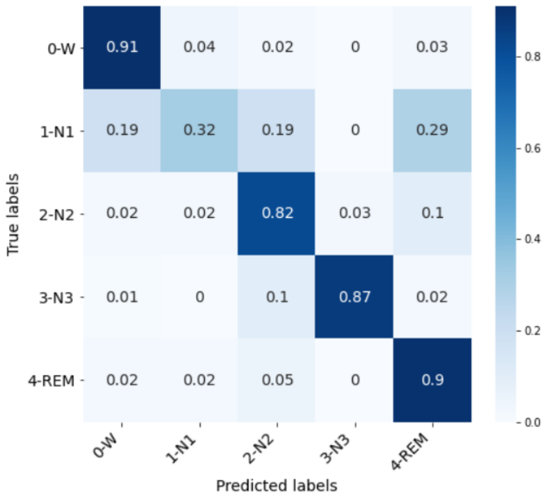

Figure 3. Confusion matrix of CNN Model

B. CNN + LSTM Model

The CNN+LSTM classifier was trained in two stages. First, the CNN used for feature extraction is trained for 19 epochs with the same parameters as mentioned in the previous section. Following that, the LSTM network is trained by using an Adam optimizer to mitigate the categorical cross entropy (lr= 0.001, 1 = 0.9, 2 = 0.99). The LSTM training is carried out in compact batches of 100 data points over 5 epochs. Figure 5 depicts the CNN + LSTM classifier's confusion matrix. The classifier achieves an MF1 score of 78.56% and an accuracy of 85.30%. The F1 score per class of the wake stage is 89.09%, N1 stage is 34.00%, N2 stage is 86.02%, N3 stage is 87.58%, N4 stage is 87.14%. The average training time was 1:18 (1 minutes and 18 seconds) for each fold, and the average prediction time for a new sleep recording was less than 1 second.

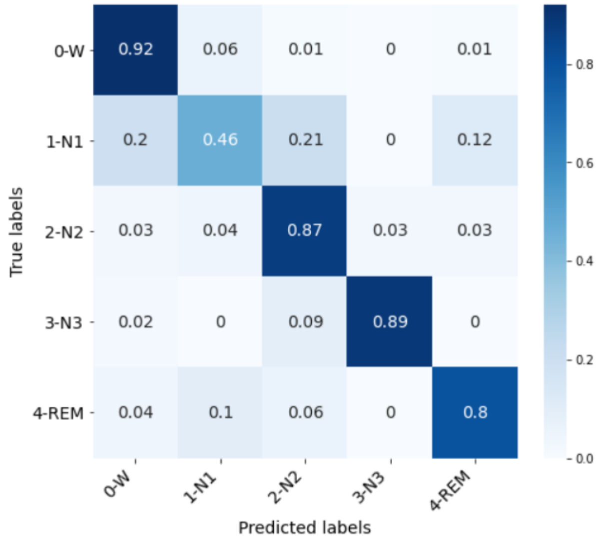

Figure 4. Confusion matrix of CNN + LSTM Model

Figure 6 shows a comparison of a real hypnogram and the classifier's predicted hypnogram to further demonstrate the accuracy of the CNN + LSTM classifier. The hypnogram depicted corresponds to a sleep recording in which the classifier achieves an accuracy similar to the classifier's overall accuracy. The hypnogram exposed is from the sleep recording SC4142E0, and the model achieved an accuracy of 85.80%, which can also be considered the percentage of similarity between both line plots. Figure 8 shows the comparison of the accuracy, MF1 score, and F1 score per class of the two above mentioned classifiers and compares the same with the existing study [9].

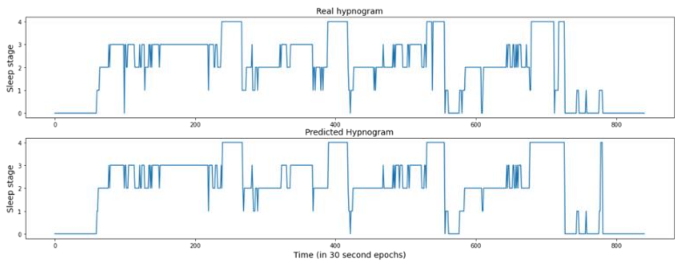

Figure 5. Real hypnogram (top) and predicted hypnogram (bottom) for CNN + LSTM classifier

TABLE I. COMPARISION OF CLASSIFIERS

Experiment	***ACC***	***MF_1***	***F_1 per class***				
			0-W	1-N1	2-N2	3-N3	4-REM
CNN	81.78	75.79	88.93	40.93	84.91	87.62	74.46
CNN + LSTM	84.19	76.76	89.09	34.00	86.02	87.58	87.14
Supratak et al.	82.00	76.90	84.70	46.60	85.90	84.80	82.40

IV. CONCLUSION

The CNN + LSTM classifier achieves the best classification performance out of the methods presented. It is noticed that, in table I, comparison with the rest of sleep stages N1 Stage is the one with the lowest F1 score. This is expected because it is the least common stage in the dataset and is not morphologically very different from other stages like the awake stage. Now, considering the hypnogram comparison presented in Figure 7, it is observed that they are similar plots. A dissimilarity found is that at the beginning of the recording the classifier predicts incorrectly REM (code 4) epochs. The fact that the classifier predicted REM stage at the beginning of the recording is not entirely rare, since it was observed in the dataset that some sleep recordings can begin transitioning directly from Wake stage to REM stage. For the remaining of the hypnogram, the plots don't differ significantly. The CNN + LSTM classifier increases the accuracy by 4 points, and reduces the duration of the task compared to the CNN model. It is noteworthy that this increase in accuracy is achieved by using a single EEG channel. It can be concluded that the present study is in par with the state of art, set in [9]. Furthermore, these results help to lay the foundations for the production of hardware for at home, automatic sleep monitoring. This paper has proven that an automatic classifier based on CNN and LSTM is capable of classifying sleep stages with a higher precision that human level sleep scoring and in significantly less time. Some of the work's limitations will be illustrated as future work. The low amount of N1 samples can be addressed by accessing a larger database or by applying data augmentation techniques.

ACKNOWLEDGEMENT

The authors would like to thank Mr Carlos Fabbri of Universidad del Pacıfico - GitHub:
https://github.com/carlosfg97/AutomaticSleepStageClassifier.

REFERENCES

[1] Salo, Paula & Polo-Kantola, Päivi.. "Sleep deprivation: Impact on cognitive performance". Neuropsychiatric disease and treatment. 3. 553-67, 2007.

[2] During, Emmanuel." The Epidemic of Sleep Deprivation: A Modern Curse", The Huffington Post, 2017.

[3] Berry, Richard & Brooks, Rita & Gamaldo, Charlene & Harding, Susan & Lloyd, Robin & Quan, Stuart & Troester, Matthew & Vaughn, Brad. "AASM Scoring Manual Updates for 2017 (Version 2.4)", Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine, 13. 10.5664/jcsm.6576, 2017

[4] Lee, M., Song, C.-B., Shin, G.-H., & Lee, S.-W, "Possible effect of binaural beat combined with autonomous sensory meridian response for inducing sleep", Frontiers in Human Neuroscience, 13, Article 425. https://doi.org/10.3389/fnhum.2019.00425 , 2019

[5] Lee M, Baird B, Gosseries O, Nieminen JO, Boly M, Postle BR, Tononi G, Lee SW, "Connectivity differences between consciousness and unconsciousness in non-rapid eye movement sleep: a TMS-EEG study", Sci Rep. 2019 Mar 26;9(1):5175. doi: 10.1038/s41598-019-41274-2. PMID: 30914674; PMCID: PMC6435892, 2019

[6] Kim, Hyeong-Jin & Lee, Minji & Lee, Seong-Whan, "End-to-End Automatic Sleep Stage Classification Using Spectral-Temporal Sleep Features", 2020. 3452-3455. 10.1109/EMBC44109.2020.9176477,2020

[7] Kemp, Bob "Model-based Monitoring of Human Sleep Stages", 1987

[8] Chambon, Stanislas & Galtier, Mathieu & Arnal, Pierrick & Wainrib, Gilles & Gramfort, Alexandre "A Deep Learning Architecture for Temporal Sleep Stage Classification Using Multivariate and Multimodal Time Series" IEEE Transactions on Neural Systems and Rehabilitation Engineering PP. 10.1109/TNSRE.2018.2813138., 2017

[9] Supratak, Akara & Dong, Hao & Wu, Chao & Guo, Yike. "DeepSleepNet: a Model for Automatic Sleep Stage Scoring based on Raw Single-Channel EEG" IEEE Transactions on Neural Systems and Rehabilitation Engineering PP. 10.1109/TNSRE.2017.2721116, 2017