Group 14: Sleep Stage Classification

Abhishek Sahu Alok Kumar Trivedi Binaya Kumar Suna
Sai Kiran Tanikella
21111002, 21111008, 21111021, 21111061
{abhisheks21, alokt21, binayas21, tskiran21}@iitk.ac.in
Indian Institute of Technology Kanpur (IIT Kanpur)

Abstract

Sleep stage classification is one of the critical methodologies for the diagnosis of sleep-related diseases and complications. The conventional method of categorization is quite clumsy and time-consuming. This project aims to devise a deep learning and machine learning model for automatic classification of sleep stage, hence, removing the barrier of conventional method and expert ubiquity. In this work, we have considered a database that carries 197-night sleep polysomnographic data. Moreover, we aimed to classify this data into stages W, N1, N2, N3 and N4 as mentioned in the AASM standard. In addition to that, we have selected the EEG FpzCz channel because of its better quality and used an epoch time of 30 seconds for signal processing. We have used four machine learning and deep learning methods, namely CNN-CNN, CNN-LSTM, Random Forest, and XGBoosting, with 82%, 87%, 51%, and 59%, respectively. This report has depicted a roadmap of the EEG-based sleep stage scoring method by implementing the state of art methods. In conclusion, using better signal processing techniques will increase the overall performance and accuracy of the model.

1 Introduction

Sleep plays an integral role in health. A good night's sleep empowers the body to recover and lets you wake up refreshed and ready to take on the day. We spend about one-third of our time on planet Earth asleep. Research has found that a lack of sleep takes a toll over both the short- and long-term, demonstrating that your brain and body can't just get used to getting less sleep. Instead, persistent sleep deprivation affects daytime performance, harming decision-making, memory, focus, and creativity. With time, insufficient sleep can wreak havoc on diverse aspects of health including metabolism, the cardiovascular system, the immune system, hormone production, and mental health.

The American Academy of Sleep Medicine has declared that Monday, March 12, 2018, Insomnia Awareness Day. Insomnia involves both a sleep disturbance and daytime symptoms. The effects of insomnia can impact nearly every aspect of your life. How common is insomnia among adults? Here are the numbers:

- 30 to 35% have brief symptoms of insomnia.
- 15 to 20% have a short-term insomnia disorder, which lasts less than three months.
- 10% have a chronic insomnia disorder, which occurs at least three times per week for at least three months.

During your sleep, your brain will oscillate between different states, also called **sleep stages** (represented here under). Each stage corresponds to particular **electric patterns** and specific brain waves. There are basically two criteria to classify the sleep pattern thoughout the night. One is Rechtschaffen and Kales i.e. R&K and the other is American academy of sleep medicine i.e. AASM.
In this paper we will be using AASM which divides its stage into

1. W: Awake state (stage W) is characterized by alpha or faster frequency bands occupying more than 50% of the epoch, frequent eye movements and high EMG tone.

2. N1: Stage N1 is scored when alpha occupies more than 50% of epoch while theta activity, slow rolling eye movements and vertex waves are evident.
3. N2: Stage N2 is scored when sleep spindles or K-complexes (less than 3 minutes apart) are noted.
4. N3: Stage N3 is characterized by delta activity detected in over 20% of the epoch length.
5. REM: Upon sleep scoring an epoch is marked as REM when saw-tooth waves along with rapid eye movements as well as lowest EMG signals are observed through each epoch.

The conventional sleep stage classification involves visual data interpretation of different EEG signals. This requires expert supervision and is a time consuming process as it takes a day to process cumulative data of 3 days EEG recording of the concerned patient. However recent development in Machine Learning has given an opportunity for automatic sleep classification. The Automatic sleep stage classification involves signal processing techniques along with efficient machine learning algorithm to obtain useful information from EEG signal to classify the sleep stages.

2 Literature Review

In this paper we have examined the different trends and methodologies of automatic sleep stage classification.

1. Multi -Class SVM based on sleep stage identification using EEG signal[1]
   Discusses the 5 different stages of sleep namely wake,stage 1,stage 2,stage 3 and stage 4. It takes the input as EEG signal in time domain and uses filtering and decomposition technique using epoch time of 60 seconds and sampling rate of 100Hz to extract the statistical features. It uses these features to train a linear kernel multiclass SVM having 80% of the dataset for training and 20% for testing. Moreover the performance estimation was done on the basis of 3 parameters namely accuracy, sensitivity and specificity and generates a total accuracy of 92%.
2. Efficient sleep stage recognition system based on EEG signal using K-means clustering based feature weighting[2]
   Discusses 3 process to classify sleep stages. The feature extraction process involve welch-spectral analysis of EEG signal using 30s epoch time. After this process we obtain 129 frequency domain features. Due to curse of dimensionality these features are reduced to 4 specific features. The data processing step uses weigting to weight the 4 extracted features and categorizes into 6 stages namely N-REM, stage 1,N-REM stage 2,N-REM stage 3, REM and movement time. The classification step involves using k-NN classifier with k value as 10,20,30,40,50 with highest accuracy of 82.21% with k value as 30.
3. Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signal[3]
   Discusses novel approach based on LSTM network for automated sleep stage classification using single-channel EEG signals. The EEG signals analysis is divided into data acquisition, signal pre-processing, feature extraction , dimensionality reduction and classification. Two RNN models were used where one was performed using multi-class classification by merging the stages N1 and REM and the other was performed using binary classification. The accuracy achieved with this model is 86.74%.
4. A study on Automatic Sleep Stage Classification Based on CNN-LSTM[4]
   Discusses a new CNN-LSTM structure for multi channel EEG signal. This process is divided into two parts where one is representation learning and the other is residual sequence learning. The residual sequence learning is further divided into 9 layer where the first layer is input layer, from second to fourth layer is spatial feature extraction layer, the fifth layer is fully connected layer, the sixth layer to the ninth layer encode time characteristics. The accuracy achieved is 84.1% which is 2% improvement over only CNN model.
5. Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier [5]
   Discusses automated sleep stage classification based on AASM standard. The feature extraction involves time-frequency analysis and time-frequency entropy measurement on a single EEG signal.

The three methods used for generating time frequency distribution are namely CWT, HHT and CWD for feature extraction. In the classification step the random forest classifier was used which gives an accuracy of 83% when CWT was used in feature extraction process. Kappa coefficient for the CWT based classification is 0.76 which shows a substantial agreement between proposed method and expert scores.

6. A new fully automated random-forest algorithm for sleep staging [6]
Discusses a sleep stage classification model based on both EEG and EOG signal recordings. The EEG and EOG was divided in overlapping moving 33s epoch with step of 3s and various time and frequency domain features were extracted. This feature was used to train random forest classifier which could efficiently label each 33s epoch with probability of wakefulness, REM and non-REM. The performance of the model was tested using 20-fold cross validation scheme where the overall accuracy achieved was 92.6% and a cohen's kappa of 0.856.

7. Automatic Sleep stage scoring with single channel EEG using Convolutional Neural Networks[7]
Discusses a sleep stage classification model based on convolutional neural network (CNN) using a single-channel EEG recording (FpzCz). A class based random sampling was used within the stochastic gradient descent (SGD) optimization to avoid skewed performance. The filters learned by the CNN was analysed and visualized and concluded that CNN learns filter that closely capture the AASM manual's guideline in terms of frequency characteristics per sleep stage. It uses 20-fold cross validation generating a mean accuracy of 82% and overall accuracy of 74%.

8. Learning a Convolutional Neural network for sleep stage Classification [8]
Discusses a sleep stage classification model based on Convolution neural network (CNN) with Electroencephalogram (EEG). The CNN model was feed a three dimensional signal consisting of EEG , fractional discrete fourier transform(F-DFT) and the wavelet transform (WT). The F-DFT and WT signal processing uses 30s epoch time achieving 90.11% accuracy.

3 Proposed Idea

We are developing Deep Learning and Machine Learning models to study EEG waves of subjects to predict sleep pattern. We have collected sleep pattern data, studied various research papers to understand the correlation between sleep hygiene and bodily EEG wave pattern, applied different signal processing technique to filter the noises present in our Dataset. We have trained model on the sleep pattern data and tries to solve a classification problem. We need to classify data into 5 different classes of sleep stage.

4 Methodology

The whole training and testing of the models are performed in Google Colab with GPU support. Below we will discuss all the steps we have done to complete the project.

4.1 Data Collection

The source and description of our Dataset used in this project is mentioned below
Source: `https://www.physionet.org/content/sleep-edfx/1.0.0/`
Description: The supplied database carries 197 whole-night times PolySomnoGraphic sleep recordings, containing EEG, EOG, chin EMG and event markers. All the documents are formatted with the "edf" extension. The ***PSG.edf** documents are whole-night time polysomnographic sleep recordings containing EEG (from Fpz-Cz and Pz-Oz electrode locations), EOG (horizontal), submental chin EMG, and an occasion marker. The ***Hypnogram.edf** documents comprise annotations of the sleep styles that correspond to the PSGs. These patterns (hypnograms) consist of sleep stages W, N1, N2, N3, N4. Well-trained technicians manually scored all hypnograms.

4.2 Data Cleaning

We have installed the MNE library to read EDF and EDF+ files. Then we fetch the data and store it in the NumPy array. It is always true that there will remain some noise in the EEG data. To remove some noise and get the best signals, we have used a band-pass filter with pass-band bandwidth of 0.5-40 Hz.

4.3 Feature Extraction

For this step, we have done the following operations

1. First, we have combined all 197 person data into one file.
2. We have selected EEG Fpz-Cz Channel from the EEG.
3. Epochs – 30s epoch used in the project.
4. We have used LabelEncoding to convert classes into Integer number.
5. Power Spectral Density was used to extract the features from filtered data. A PSD is typically used for random broadband signals.
6. In 1 sec, a total of 100 recordings were extracted. So, in 1 epoch, a total of 3000 recordings were gathered.
7. Now for every epoch, we convert the list of 3000 recordings into features. So from linear vector, now we have a matrix where each row has 3000 features.

4.4 Splitting Data

For any Machine Learning/ Deep Learning Model, we need to have the training and testing data. For this, we have split the Dataset into training and testing data with a ratio of **80:20**.

4.5 Class Balancing

The Dataset that we are using has an imbalance in the classification of sleep patterns, where one class has more data than another class. So to have a better prediction, we need to balance the Dataset. For doing this step, we have used a technique called "SMOTE".
SMOTE works by selecting close examples in the feature space by drawing a line between the standards in the feature space and drawing new samples at a point along that line.
This technique helps us to balance all the classes with an equal number of classes. And after class balancing, we have again split the training Dataset into training data and validation data with a ratio of **90:10**.

4.6 Predictive Models

We deployed CNN-CNN and CNN-LSTM deep learning models and various machine learning models such as Random Forest, XGB classifier for sleep stage classification with EEG time series data.

In our predictive model a 30 second long EEG segment is taken as a single Dataset and target variable consist of 5 integers which indicates the sleep pattern of the subject in that 30 second long segment. In subsequent paragraphs, we will present our Deep Learning and Machine Learning architecture .

1) CNN-CNN Architecture: In order to efficiently extract the features and transform the biological waves to numbers , we included two CNN architecture in our model. we designed base model ,which can be used for both CNN-CNN and CNN-LSTM model with some minor changes . It consist of one input layer, six conv1D layers,three maxpool1D layer and 3 spatial dropout layer and one dense layer, while final CNN layer has one input layer, one Time distributed layer,one dropout and one conv1d layer.

we have briefly described each layer used in the CNN-CNN model

- convolution 1D layer - Unlike convolution 2D , here convolution kernel convolve with the input layer in one dimension and generated output is also in one dimension.

- Max pool 1D layer - Max pool 1D layer is responsible for downsampling the input data by choosing the max value in a given patch size along a dimension.
- SpatialDropout1D layer - It drops entire feature map along a dimension with a given probability
- Global Max pool 1D layer - It downsamples the input vector by taking maximum over time dimension
- Dense layer - It is also called "Fully connected laye", It is usually last few layer in the CNN architecture.It connectes each input node to output node.
- TimeDistributed layer - It acts like a wrapper which allow to apply a layer to every input data.

Details of Base CNN model

layer	number	type	Kernel size
Input			
convo1D	16	Relu	10
convo1D	16	Relu	10
Maxpool1D			3
SpatialDropout1D			prob = 0.01
convo1D	32	Relu	5
convo1D	32	Relu	5
Maxpool1D			2
SpatialDropout1D			prob = 0.01
convo1D	128	Relu	3
convo1D	128	Relu	3
GlobalMaxPool1D			
Dropout			prob = 0.01
Dense	64	sigmoid	
Dropout			prob = 0.01

Detail of final CNN architecture

layer	number	type	Kernel size	Output
Input				(None, None, 3000, 1)
Time distributed				(None, None, 64)
convo1D	256	Relu	6	(None, None, 256)
Dropout			0.01	(None, None, 256)
convo1D	5	sigmoid	3	(None, None, 5)

As shown in above tables,we have encapsulated two CNN architecture in our model. Firstly we reshape our input vector to suit the model then we apply two consecutive convolution1D layes with kernal size 10 followed by one maxpool layer of size 3 and one dropout layer . This conv-conv-maxpool-dropout unit is applied three times on the dataset and then the output is fed to the dense layer which uses sigmoid function to product most probable output. The output from the dense layer is then fed to dropout layer to dropout layer to prevent the model from overfitting

The base model is then fed to a time distributed layer, which acts as a wrapper responsible for applying base model to every input sequence of the Dataset. we then apply Convolution 1D layer with Relu activation function to extract the feature from the input data ,followed by a dropout layer to curb the overfitting ,we then apply another conv1d layer with sigmoid activation function to get the most probable label as the output .

The model is compiled with **Adam Optimizer** with loss function **Sparse Categorical Cross Entropy loss function** and metrics as **Accuracy**. To train our model we have used **epoch=**500. Since it may cause overfitting, so we have used a method called **Early-Stopping**. It will monitor validation loss and stop training the model if after somepoint the validation loss increases.

2) CNN-LSTM Architecture: We have proposed a CNN architecture coupled with LSTM model to effectively predict the sleep pattern of a subject. The base model is same as the base model of

the CNN-CNN architecture with different hyperparameter values . The LSTM architecture of model consist of one input layer,one time distributed layer, one bidirectional layer, 2 dropout layer and a final conv1D layer with sigmoid activation function which gives us the final output.

Detail of final LSTM architecture

layer	number	type	Kernel size	Output
Input				(None, None, 3000, 1)
Time distributed				(None, None, 64)
Bi-directional	100			(None, None, 200)
Dropout			0.1	(None, None, 200)
Dropout			.07	(None, None, 200)
conv1D	nlabel	sigmoid	1	(None, None, 5)

In this section we discuss working of LSTM model . The input is fed to the time distributed layer which acts as a wrapper and allow the base layer to be applied to every input sequence, the output of time distributed layer is then fed to a bi-directional layer which makes the input data to flow in both forward and backward direction for effective feature extraction. After that two consecutive dropout layer is applied to prevent model from overfitting and then we get the most probable label as output by passing the sequence of data to a conv1D layer.

The model is compiled with **Adam Optimizer** with loss function **Sparse Categorical Cross Entropy loss function** and metrics as **Accuracy**. To train our model we have used **epoch=**500. Since it may cause overfitting, so we have used a method called **Early-Stopping**. It will monitor validation loss and stop training the model if after somepoint the validation loss increases.

3) Random Forest: We have also deployed Random Forest for sleep pattern prediction. Random Forest Architecture - We utilised 100 decision trees with max depth = 10.

4) XGBoosting Classifier: Another ensemble method used for sleep pattern prediction.

5 Results

Performance metric- In order to measure performance of our developed, we have defined certain parameters.

- Accuracy - It is a ratio of umber of correctly predicted label to total observations present in the Dataset

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

- Precision - It is a ratio of correctly predicted number of a label to total number of predicted value of a label

$$Precision = \frac{TP}{TP+FP}$$

- Recall - Recall is the ratio of total correctly predicted true positive to total positive label present in the

$$Recall = \frac{TP}{TP+FN}$$

- f1 score - f1 score is Harmonic mean of Precision and Recall. It takes false negative and false positive in account while judging the performance of a model. It works better than accuracy in a extremely imbalanced data set.

$$f1score = \frac{2*precison*recall}{precison+recall}$$

Evaluation: In this section we will discuss the performance of our models and compare these two model in terms of accuracy and f1 score.

Evaluation Report for CNN-CNN Model

label	precision	recall	f1 score
0	.92	.86	.89
1	.72	.74	.73
2	.80	.77	.79
3	.90	.95	.92
4	.78	.79	.79
accuracy			.82
macro avg	.82	.82	.82
weighted avg	.82	.82	.82

Evaluation Report for CNN-LSTM Model

label	precision	recall	f1 score
0	.95	.89	.92
1	.78	.82	.80
2	.87	.83	.85
3	.93	.97	.95
4	.84	.85	.84
accuracy			.87
macro avg	.87	.87	.87
weighted avg	.87	.87	.87

Evaluation Report for Random Forest Model

label	precision	recall	f1 score
0	.77	.34	.47
1	.46	.71	.56
2	.46	.48	.47
3	.68	.99	.81
4	.38	.20	.26
accuracy			.54
macro avg	.55	.54	.51
weighted avg	.55	.54	.51

Evaluation Report for XGBoosting Model

label	precision	recall	f1 score
0	.76	.39	.52
1	.52	.66	.58
2	.52	.52	.52
3	.74	.93	.82
4	.45	.43	.44
accuracy			.59
macro avg	.60	.59	.58
weighted avg	.60	.59	.58

CNN-CNN has achieved accuracy of 82 percent with a f1 score of 82 percent. It has achieved a accuracy of above .9 for label '0' and '3' and label'2' has .8 accuracy.

Our CNN-LSTM model has performed relatively better with accuracy of .87 and f1 score as .87. it has achieved accuracy over 0.9 for label '0' and '3' and every other label has accuracy above or around 0.8.

XGBoost also an light weight model with accuracy pf 59 percent and Random Forest with accuracy of 51 percent.

6 Related Work

There has been some similar work done about sleep classification. From that previous work, we have found that some have used Deep Learning models like LSTM and CNN, and some have used machine learning models like Random Forest, XGBoosting. In all these papers, they have trained in well-defined Datasets. In our Project, we have worked on entirely raw data. Our primary step involves the data preprocessing part. We have given more focus to the feature extraction part. And we have implemented the models without any use of the transfer learning technique. Our models, i.e. CNN-CNN and CNN-LSTM model, suits the best as the first step of these model help use for data preprocessing and the second part help to predict. Again in the sleep dataset, it's an entirely sequential database. So use of memory-based technique help for better predictions.

7 Discussion and Future Work

Our model has performed reasonbly well in predicting the sleep pattern of the subject. We will try to improve the accuracy of label '1' (N1 stage of sleep) which has significantly lower accuracy than other label. we We will learn more about signal processing to carry out better pre-processing so that we can remove some noises still present in the data. we would also generate EEG data to test our model against it.

8 Conclusion

We have developed two deep learning model to predict the sleep hygiene of subject using EEG waves as dataset . Our models has successfully predicted the sleep pattern with .82 and .87 accuracy. We have experienced different accuracy for different sleep label. We have found that label '1' which indicates N1 stage of sleep(the nascent stage of sleep) has lowest accuracy among all other labels.We found that low accuracy could be the result of noise present in the dataset which did not get filtered out during the data preprocessing step and due to unaccounted external factors which disturbed the sleep of the subjects while conducting the experiment such as sudden change in temperature and humidity or some loud noises.

These unaccounted external factors are more likely to affect the sleep of the subject when the subject has just made a transition from wake up state('0) to first stage of sleep 'N1'('1'). Thus these factors were able to affect the accuracy of label '1' and the overall accuracy of the model.

We also deployed machine learning model but we did not get as high accuracy from these model as expected. Sleep pattern Datasets are usually very large and complex involves very intricate details that need to be extracted as features. Machine learning models are generally unsuccessful in extracting these features.

9 Individual Contributions

Abhishek: Data Collection: 40%, Preprocessing: 30%, Model: CNN-LSTM model, Report: 30%
Alok: Literature Review: 30%, Data Collection: 40%, Model: CNN-CNN model, Report: 30%
Binay: Literature Review: 40%, Preprocessing: 40%, Model: XGBoost model, Report: 20%
Kiran: Literature Review: 30%, Data Collection: 20%, Preprocessing: 30%, Model: Random Forest model, Report: 20%

References

[1] K. A. Aboalayon and M. Faezipour, "Multi-class svm based on sleep stage identification using eeg signal," in *2014 IEEE Healthcare Innovation Conference (HIC)*, pp. 181–184, IEEE, 2014.

[2] S. Güneş, K. Polat, and Ş. Yosunkaya, "Efficient sleep stage recognition system based on eeg signal using k-means clustering based feature weighting," *Expert Systems with Applications*, vol. 37, no. 12, pp. 7922–7928, 2010.

[3] N. Michielli, U. R. Acharya, and F. Molinari, "Cascaded lstm recurrent neural network for automated sleep stage classification using single-channel eeg signals," *Computers in biology and medicine*, vol. 106, pp. 71–81, 2019.

[4] Y. Yang, X. Zheng, and F. Yuan, "A study on automatic sleep stage classification based on cnn-lstm," in *Proceedings of the 3rd International Conference on Crowd Science and Engineering*, pp. 1–5, 2018.

[5] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, "Automated sleep stage identification system based on time–frequency analysis of a single eeg channel and random forest classifier," *Computer methods and programs in biomedicine*, vol. 108, no. 1, pp. 10–19, 2012.

[6] A. B. Klok, J. Edin, M. Cesari, A. N. Olesen, P. Jennum, and H. B. Sorensen, "A new fully automated random-forest algorithm for sleep staging," in *2018 40th annual international conference of the ieee engineering in medicine and biology society (EMBC)*, pp. 4920–4923, IEEE, 2018.

[7] O. Tsinalis, P. M. Matthews, Y. Guo, and S. Zafeiriou, "Automatic sleep stage scoring with single-channel eeg using convolutional neural networks," *arXiv preprint arXiv:1610.01683*, 2016.

[8] N. Liu, Z. Lu, B. Xu, and Q. Liao, "Learning a convolutional neural network for sleep stage classification," in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1–6, IEEE, 2017.