Automated phase classification in cyclic alternating patterns using convolutional recurrent neural networks.

ARTICLE INFO

Keywords:
Cyclic alternating patterns
Electroencephalogram (EEG)
Inception modules
Gated recurrent units (GRU)
k-fold stratified cross validation

ABSTRACT

Sleep is one of the important parts of a person's daily life, the dissatisfaction of which can lead to a disrupted lifestyle. The detection of microstructure such as cyclic alternating patterns (CAP) in EEG signals is a crucial factor in the diagnosis of sleep which can help in discerning sleep disorders such as sleep apnea, insomnia, etc. However, the sleep scoring methods used by experts today are tedious and prone to errors. These diagnostics can be accelerated without compromising on accuracy using artificial intelligence. In this paper, we propose a deep learning model based on Inception modules paired with gated recurrent units (GRU) for the classification of CAP phases (A & B). Inception modules help in the feature extraction without increasing the load on computational resources while GRU help in the sequential processing of data. Our model is tested with various validation methods and can achieve an accuracy of 74.64% on a balanced dataset consisting of data of 6 subjects, further tested on a dataset extended to 14 subjects achieving an accuracy of 68.07%. The proposed model is ready to be tested with more data and can be deployed to aid somnologists in assessing the patient's cerebral activities and detect any underlying diseases.

1. Introduction

Sleep is a vital activity that plays a crucial role in the physical and mental health of the human body. A healthy sleep helps the brain to function properly and repulse away diseases [1]. Research has proven that teenagers require 8-10 hours of sleep while adults below the age of 60 need 7-9 hours with the rest requiring even lesser hours [2, 3]. Missing out on sleep can affect one's lifestyle by causing irritation, mood deviations, inability in decision-making, and problem-solving [3]. As we lead a fast-paced lifestyle, we fail to understand the importance of sleep which can disrupt our daily life and yet most of the sleep disorders are untreated in clinical practice. Such disorders affect around 50 to 70 million people in the United States of America (USA) [4]. According to the World Health Organization (WHO) and the International Network for the Demographic Evaluation of Populations and their Health (IN-DEPTH), about 16.6% of the adult population in underdeveloped nations suffer from sleep disorders [4]. In the USA, 30% of individuals suffer from insomnia, a condition that makes it hard to fall asleep [4]. Individuals associated with insomnia and obstructive sleep apnea (OSA) are sensitive to a variety of health issues including strokes, cardiovascular diseases, and obesity. According to a study conducted by [5], around 33% of the population of India suffers from insomnia. OSA is one of the most common sleep disorders. Reports have also stated that more than 2% of the female and 4% of the male population has been diagnosed with sleep disorders [6].

Polysomnogram (PSG) recordings of subjects are the physiological signals that are collected overnight for sleep analysis and scoring. PSG is a multivariate signal, i.e., it consists of several distinguishable components including electroencephalogram (EEG), electrocardiogram (ECG), electrooculogram (EOG) and electromyography (EMG) signals [7, 8]. Typically, a sleep expert evaluates a PSG signal within a specific time frame, usually 30 seconds in duration, and then determines the sleep score based on various criteria.

A sleep cycle comprises of recurring alternating patterns followed by rapid eye movement [1]. To discern its macrostructure, based on Rechtschaffen and Kale's (R&K) rules [9, 10], the sleep stages are classified as wakefulness (W), rapid eye movement (REM), and non-rapid eye movement (NREM) which consists of 4 stages (S1-S4) [11]. Subsequently, the American Academy of Sleep Medicine (AASM) updated these stages and coalesce S3-S4 into one stage known as slow wave sleep (SWS) due to similar characteristics evolved in five stages as described below [12]:

- Wakefulness (W): The brain is most active in this stage. EEG signals in this stage comprises of high alpha rhythms and occasional beta rhythms [13].
- Non Rapid Eye Movement (NREM) Stage 1: This is referred as the light sleep stage. Individuals tend to wake up easily at this stage by disruptive noises. The muscles are relaxed, the heart rate slows, and the body temperature drops. Alpha rhythms vanish, and theta signals start appearing [13].
- NREM Stage 2: An individual's eye movements cease at this stage and respiration slows down. Sleep spindles and K-complexes occur for approximately 1-2 seconds [13].
- NREM Stage 3 & 4: SWS is the deepest phase of NREM sleep which results in reduced blood pressure and restored muscle growth. The brain flushes out the unnecessary fluids and excess protein build ups, resulting in long, slow brain waves. An individual finds it difficult to wake up at this stage and is disoriented and groggy if they do. Low-frequency theta signals start to appear [13].
- Rapid Eye Movement (REM): Individuals experience vivid dreams at this stage. The heart rate rises and the

ORCID(s):

temperature regulators turn off. The body becomes immobile, and control over the body is lost. This stage is composed of theta rhythms and the flattening of EEG signals [13].

EEG signals not only provide macrostructural information for sleep stages like NREM, REM and wakefulness but also microstructural features such as cyclic alternating patterns (CAP) [1, 14]. CAP is a periodic EEG activity that represents sustained arousal instability alternating between higher and lower arousal levels classified as phase A and phase B, respectively [11]. All CAP cycles begin with A phases and are followed by B phases, each of them lasting for a duration of 2-60 seconds [15]. They are critical in detecting sleep disorders such as insomnia and narcolepsy. It includes sleep events such as falling asleep, changing sleep stages, and waking up [16]. If CAP occurs during the REM stage, it is classified as a sleep disorder.

Typically, phase B refers to the background activities occurring in the signal. Brain activities, including cortical arousal, are depicted in phase B. Phase A consists of slowly varying high amplitude or quickly varying low amplitude, and can last for 2-60 seconds in the NREM stage [16]. Phase A can be subdivided into three subtypes: A1, A2, and A3 [8, 11].

- Subtype A1: These waves have high amplitude and low frequency. The pattern must occupy 80% of the sample to be classified as A1. The frequency ranges from 0.5-4 Hz [17], and it's designated by delta bursts and K-complexes [8].
- Subtype A2: These waves are a subset of subtypes A1 and A3. They are a mixture of low-frequency and high-frequency signals [8].
- Subtype A3: These waves are of low amplitude and high frequency. K-alpha, EEG arousal, and polyphasic bursts are the waveforms associated with this subtype. A3 consists of alpha waves of frequency 8-12 Hz and beta waves of frequency 12-30 Hz [17, 8].

EEG signals can be subdivided into five categories based on their frequency as - delta, theta, alpha, beta, and gamma waves [18]. The various sleep stages are characterized by the spectral content of these neural oscillations. Delta waves range from 0.5-3 Hz with a high amplitude and are aroused when a person is in a deep sleep [19]. Theta waves ranges from 4-7 Hz and usually represent the various sleep stages. Alpha waves range from 8-13 Hz, depicting the relaxed stages of the sleep. Beta waves range from 14-30 Hz, and are aroused during the thinking stages [20]. Gamma waves on the other hand range from 31-60 Hz, and are associated with REM.

Feature extraction, feature selection and classification steps are generally used for the processing of EEG signals. Time, frequency, and time-frequency domain-based transformations are applied for feature extraction [17]. Machine learning (ML) and deep learning (DL) algorithms are preferred for classification and prediction purposes as the complexity of the signals and their processing advances. The architecture can be designed as shallow or deep, depending on the complexity of the signal [21]. Deep architectures tend to extract more features and can lead to some remarkable results, although they require a lot of training data and time.

In the present day, CAP cycles are analyzed to detect sleep disorders. The AASM advises that any suspected sleep disorder being analyzed using CAP cycles, should be done by consulting a certified professional.

Sleep experts score the sleep stages manually which can be time-consuming and labor-intensive. Detecting minute and rapid changes in EEG signals is imperceptible to humans, even experts, making the process prone to errors [8]. Sleep lab diagnoses are very expensive, and patients are required to stay overnight under observation [21]. Such sleep labs are only found in major cities such as the Division of Sleep Medicine at Stanford University, European Sleep Research Society etc. [22, 23]. DL has the potential to ease this process by a huge scale and also reduces the need of expertise to interpret the results. DL techniques can be used to create tools that aid in determining the sleep stages of normal and disordered patients.

Researchers make use of various methods for the classification of CAP and usually deploy them on a common dataset to achieve the best results. One such approach is the K-nearest neighbors (KNN) classifier that locates the scattered classes and assigns labels to the common pattern. [24] achieved an accuracy of 80% for detecting phase A signal with a sampling frequency of 1000 Hz on an unbalanced dataset. Another such ML based classifier is the support vector machine (SVM) which minimizes the error instead of minimizing classification errors. [25] achieved an accuracy of 84% on an unbalanced dataset consisting of signals sampled at 100 Hz. An automated technique based on the synthesis of five descriptors from the EEG signals sampled at 100 Hz and capable of providing meaningful data reduction has been also been created. This method attains an accuracy of 87% on the unbalanced dataset. Variable long short-term memory (LSTM) quantifies temporal EEG behavior, resulting in more precise scoring of conventional sleep stages. This approach was implemented on both balanced and unbalanced datasets, resulting in an accuracy of 82.42% [1].

Recently, the existing data in the Physionet CAP Sleep database [26, 27] was sampled at 512 Hz. These new samples were used by [28] in their work on Wigner-Ville based entropy feature extraction, which resulted in an accuracy of 72.35% on a balanced dataset consisting of 9306 samples. The one dimensional convolutional neural networks (1D-CNN) designed by [10] for the detection of A and B phase cyclic alternating signals in the same balanced dataset consisting of 9306 samples has resulted in an accuracy of 73.6%. This CNN is also able to achieve an accuracy of 90.4% on 3 stage sleep classification.

Our contribution: In this paper, we propose a custom deep learning model which hybridizes 1D-CNN and recurrent neural networks (RNN), also known as convolutional recurrent neural network (CRNN) for the automated classi-

fication of phases in cyclic alternating patterns. Some major highlights of our work include:

- A first of its kind neural network architecture which hybridizes dimensionality reducing Inception modules with gated recurrent units (GRU).
- We have tested our network on two sample sizes; an initial dataset with 6 subjects with consistent sampling frequency and an extended dataset with 14 subjects.
- To ensure robust training and reduce bias to a particular class, we have used 15% hold-out validation (HoV), 5-fold and 10-fold stratified cross validation (CV) methods to evaluate our model.

The details of the acquired data are elaborated in Section 2.1, and experimental setup in Section 2.2. The proposed architecture along with the major modules used are explained in 2.3. The results are covered in Section 3.

2. Methodology

2.1. Data Preparation

Publicly available CAP-sleep database has been utilized for this experiment [26, 27]. The database consists of 108 PSG recordings taken at the Sleep Disorders Center at the Ospedale Maggiore in Parma, Italy [29]. The CAP database contains recordings of 16 healthy participants which includes 15 EEG channels, 2 EOG channels, 2 EMG channels, as well as ECG and respiration signals. The data of the 16 normal patients present in the database is sampled between 128 - 1000 Hz [10].

Subject	A	B	Total
n1*	1062	1062	2124
n2*	981	981	1962
n3*	852	852	1704
n4	342	342	684
n5*	471	471	942
n6	1389	1389	2778
n7	684	684	1368
n8	426	426	852
n9	528	528	1056
n10*	477	477	954
n11*	810	810	1620
n12	207	207	414
n15	1482	1482	2964
n16	1281	1281	2562
Total	**10992**	**10992**	**21984**

* subjects used in initial dataset with consistent sampling frequency.

Table 1: Number of samples acquired and balanced for individual subjects.

In this study, the data utilised consists of either C4-A1 or C3-A2 of EEG channels. For initial works, only six subjects namely n1, n2, n3, n5, n10 & n11 were considered to maintain the consistency in sampling frequency. Later, the dataset was extended which consisted the data of fourteen subjects n1-n12, n15 and n16, increasing the size of the dataset. To ensure consistency, the subjects with originally a different sampling frequency were re-sampled to 512 Hz.

A deep learning model needs to be trained robustly with equal number of samples for every class. The data acquired was unbalanced with unequal number of samples of A and B phases. Hence the unbalanced dataset was balanced with the help of random selection of samples of each class. The number of samples acquired for each subject are noted in Table 1. A total of 9306 samples were acquired post-balancing for the initial dataset. On extending the dataset, the sample size increased to 21984.

2.2. Experimental Setup

The balanced dataset was preprocessed by the standardization of the EEG samples. A discrete Fourier transform (DFT) was applied and followed by taking its magnitude spectrum. Due to the symmetric property of DFT, only half of the spectrum was considered, thereby reducing computation cost in our deep learning model. After applying DFT and discarding the symmetric values, a MinMaxScaler function was used to scale all the magnitude spectra to a range of 0 to 1 using the global minimum and maximum values in the dataset to ensure a uniform scaling. The scaled values were fed as the input to the neural network.

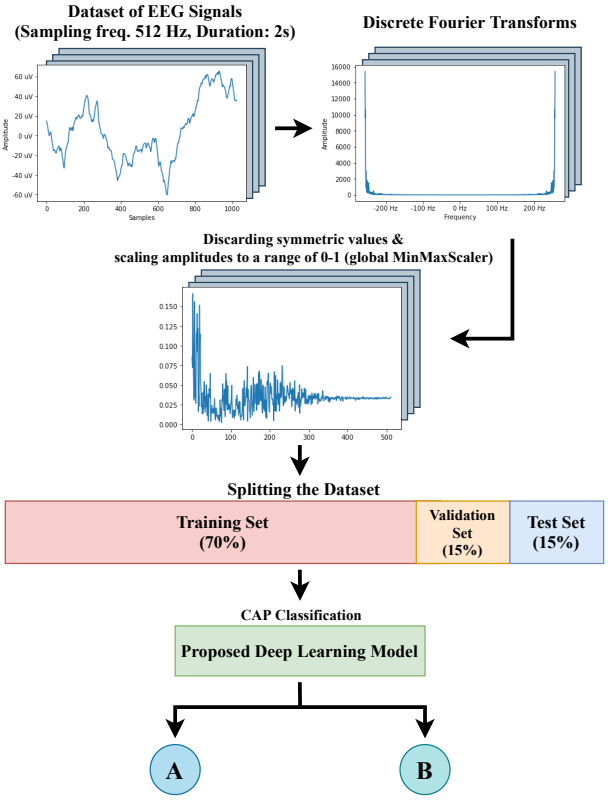

Figure 1: Workflow of the proposed method.

The dataset was split into a training set (70%), valida-

tion set (15%), and test set (15%) for the hold-out validation (HoV) method. The workflow of this experiment can be visualized in Fig. 1.

With numerous trials and errors, the hyperparameters for the model were finalized and are summarized in Table 3. The model was also subjected to a stratified 5-fold and 10-fold cross-validation (CV) method to estimate the abilities of our deep learning model. Compared to the conventional k-fold CV, the stratified CV method ensures that the train-test split consists of an equal number of A and B samples such that the model trains robustly and without any bias to a certain class.

2.3. Proposed Deep Learning model

The presented convolutional recurrent deep learning model can be divided into two parts: the inception block and the gated recurrent unit (GRU) layers. These two parts along with two other convolutional layers between them form up one convolutional recurrent (CR) block as depicted in Fig. 2. Five such CR blocks have been used in our architecture.

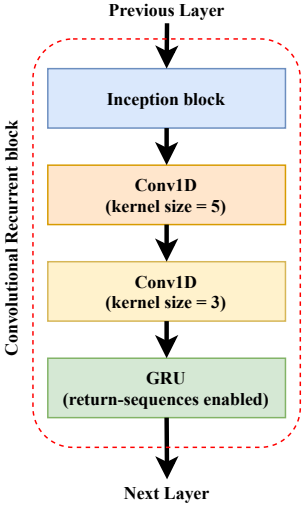

Figure 2: Convolutional Recurrent Block used in our proposed architecture.

2.3.1. Inception modules

Motivated from the GoogLeNet architecture also known as Inception v1, developed by C. Szegedy et al [30] for their submission in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014. GoogLeNet is a 22 layer deep CNN recognized for its remarkable results in computer vision tasks such as facial recognition, object detection, adversarial training, etc. It was designed with the idea to minimize computational costs and at the same time build a deep, wide architecture. It considers multiple convolutional layers of different kernel sizes at the same level, rather than stacking them sequentially. We have utilized the Inception module with dimension reductions which is based on the success of embeddings. The low dimensional embeddings might contain abundant information about a relatively large region, which is extracted by the 1×1 convolutional layers, before the expensive 3×3 and 5×5 convolutional layers come into act.

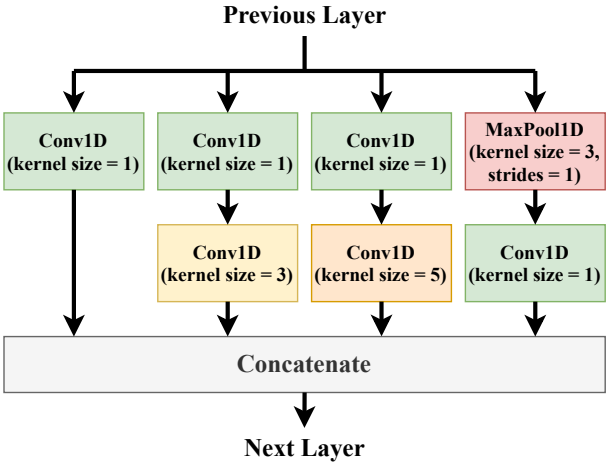

Figure 3: Inception block implemented using 1D convolutions, derived from the famous GoogLeNet (Inception v1) architecture [30].

The Inception v1 architecture was designed for RGB images, i.e a 3D input. Our version of the Inception module uses 1D convolutions instead of 2D to ensure compatibility of the input tensors (each input sample has the shape of 512, 1). The block has been diagrammatically represented in Fig. 3.

Convolution is the most important operation involved in our neural network. It extracts features from the input by computing over a small region using kernels of a predefined size. Most often, a kernel size of 1×1, 3×3, and 5×5 is preferred, whose values, also known as weights, are the parameters that are learned by the deep learning model. These weights are constantly updated throughout the training process such that the loss is minimized and the model makes accurate predictions. The Keras API used for designing the network, allowed us to set the various parameters of the convolutional layer. The number of filters were added in a decreasing order along with a ReLU activation function and same padding such that no features are lost from the edges during the convolution operation. Apart from the CR block that forms the basis of our architecture, max-pooling layers were also added between every block to downsample the features. The pool size and strides were set to the default value 2.

2.3.2. Gated recurrent units

Initially introduced by Rumelhart et al. [31], most RNNs can process temporal sequences of variable length by forming a directed graph using the connections between the neurons, exhibiting a temporal dynamic behavior. The most common form of RNN is the LSTM, which was used by Hartmann et al. [1] for their A-phase detection model.

In the recurrent segment of our architecture, we have used gated recurrent units (GRU) which were first introduced

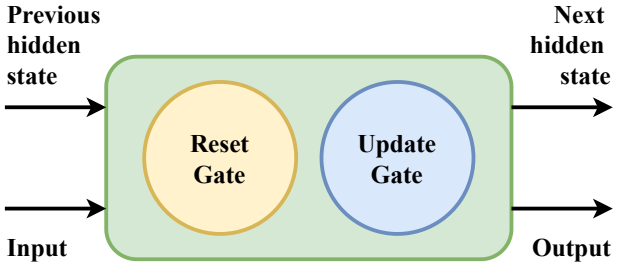

Figure 4: A gated recurrent unit.

by Cho et al. [32] in 2014. GRU is similar to LSTM but has fewer parameters as it has only two gates: reset and update, compared to LSTM which has three gates: input, output, and forget. This results in a reduced computational cost, and yet the performance of GRU is at par with LSTM, sometimes even outperforming them in certain tasks when smaller and less frequent datasets are used [33]. A GRU block is shown in Fig. 4. In our architecture, the unit size of the GRU layers have been set in decreasing order with the return sequences parameter enabled such that the full sequence is returned to the output sequence [34].

2.3.3. Classification

After the features of the input were extracted and learned using the CR blocks, the resulting output was fed to a flatten layer that converts the input into a single column tensor that can be passed on to the fully connected layer. Fully connected layers, or the Dense layer, compiles the features extracted by previous layers to form the final output. In our model, we have used a Softmax activated dense layer that gives the probability distribution over the prediction of the two output classes. The model architecture is shown in Fig. 5 and the layer parameters are summarized in Table 2.

The sliced magnitude spectra of the DFT are fed as the input to the model in a batch size of 64. A larger batch size can lead to faster training but poor generalization, whereas a small batch size can help in better convergence [35]. The CR blocks have been designed in such a way that the channel size reduces to half after passing through them, which can be noted from the Table 5. The max-pooling layer reduces the input size by half. After passing through 5 CR blocks and 4 max-pooling layers, the input shape reduces from $(512, 1)$ to $(32, 8)$. Adam optimizer was used for training to update weights and minimize the loss of the model, with a learning rate of 0.0001. A binary crossentropy loss function was used since we require a binary classification (A and B) and is formulated by:

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \log \hat{y}_i + (1 - y_i).\log(1 - \hat{y}_i)$$

The hyperparameter settings used in this experiment are summarized in Table 3. The model was set to train for 2500 epochs with an early stopping callback monitoring the validation loss such that the training process stops when the

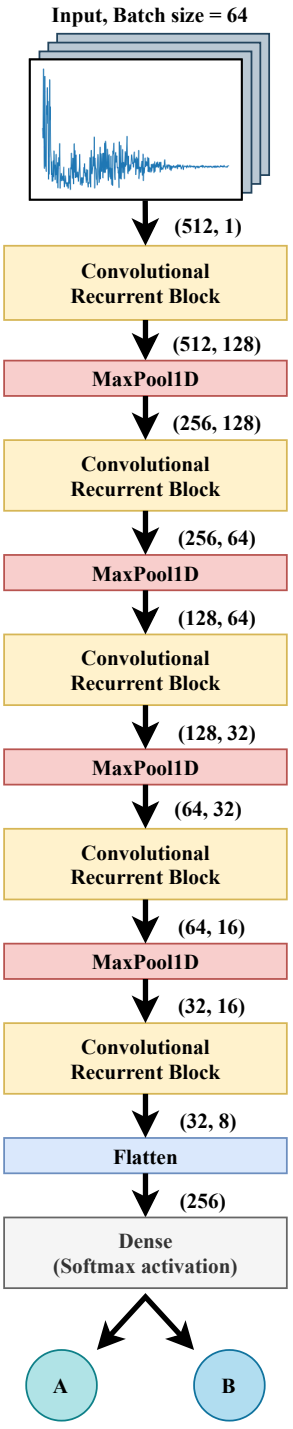

Figure 5: Architecture of the proposed deep learning model.

model attains a minimum validation loss, with the patience of 20 epochs. If a new minimum value for validation loss occurs during the patience epochs, the callback resets and the training process continues. The callback also restores the best weights of the model at the end of the training, i.e the weights from the epoch where the model reaches minimum validation loss. The validation split parameter sets aside a portion of the training dataset for validation purposes. We

#	Layer	No. of Filters Kernel size=1	No. of Filters Kernel size = 3	No. of Filters Kernel size = 5	Units	Layers Settings	Output Shape	No. of Trainable Parameters
1	Input	-	-	-	-	-	512, 1	0
2	Inception block	32	64	128	-	ReLU activation, Same padding	512, 256	27072
3	Conv1D	-	-	192	-	ReLU activation, Same Padding	512, 192	245952
4	Conv1D	-	128	-	-	ReLU activation, Same Padding	512, 128	73856
5	GRU	-	-	-	128	Return Sequences enabled	512, 128	99072
6	MaxPool1D	-	-	-	-	Pool size = 2, Stride =2	256, 128	0
7	Inception block	16	32	64	16	ReLU activation, Same padding	256, 128	15008
8	Conv1D	-	-	96	-	ReLU activation, Same Padding	256, 96	61536
9	Conv1D	-	64	-	-	ReLU activation, Same Padding	256, 64	18496
10	GRU	-	-	-	64	Return Sequences enabled	256, 64	24960
11	MaxPool1D	-	-	-	-	Pool size = 2, Stride =2	128, 64	0
12	Inception block	8	16	32	8	ReLU activation, Same padding	128, 64	3792
13	Conv1D	-	-	48	-	ReLU activation, Same Padding	128, 48	15408
14	Conv1D	-	32	-	-	ReLU activation, Same Padding	128, 32	4640
15	GRU	-	-	-	32	Return Sequences enabled	128, 32	6336
16	MaxPool1D	-	-	-	-	Pool size = 2, Stride =2	64, 32	0
17	Inception block	4	8	16	4	ReLU activation, Same padding	64, 32	968
18	Conv1D	-	-	24	-	ReLU activation, Same Padding	64, 24	3864
19	Conv1D	-	16	-	-	ReLU activation, Same Padding	64, 16	1168
20	GRU	-	-	-	16	Return Sequences enabled	64, 16	1632
21	MaxPool1D	-	-	-	-	Pool size = 2, Stride =2	32, 16	0
22	Inception block	2	4	8	2	ReLU activation, Same padding	32, 16	252
23	Conv1D	-	-	12	-	ReLU activation, Same Padding	32, 12	972
24	Conv1D	-	8	-	-	ReLU activation, Same Padding	32, 8	296
25	GRU	-	-	-	8	Return Sequences enabled	32, 8	432
26	Flatten	-	-	-	-	-	256	0
27	Dense	-	-	-	2	Softmax activation	2	514
Total no. of Trainable parameters = 606226								

Table 2: Proposed model architecture

have set the validation split as 18% (15% of the total dataset) for the hold-out method and 20% for the 5-fold and 10-fold cross-validation method.

Hyperparameters	
Epochs	2500
Batch Size	64
Optimizer	Adam
Loss function	Binary crossentropy
Learning Rate	0.0001
Validation Split	0.18
Callbacks	Early Stopping (patience = 20)

Table 3: Hyperparameters for the proposed deep learning model.

Various performance parameters apart from accuracy were noted for evaluation for the model. The results have been discussed and compared with previous works in section 3.

3. Results and Discussions

This section reports the various parameters that were used to measure the performance of the model. Along with the various accuracies, we also noted the precision, sensitivity, specificity and F1-score.

The proposed deep learning model was coded using Python 3.7 and Keras API, and trained in Google's Colaboratory platform, accelerated with a Nvidia Tesla K80 GPU and 12 GB RAM. GPU acceleration helped the model to train using hold-out method under 15 minutes, on an average of 8 seconds per epoch. The time taken to complete training for every validation method was noted in Table 4.

Method	Time taken (hh:mm:ss)	
	Initial dataset	Extended dataset
15% HoV	00:12:25	00:26:49
5-fold CV	01:25:15	02:56:09
10-fold CV	03:06:58	06:47:13

Table 4: Time taken for each validation methods to complete training.

3.1. Initial dataset with 6 subjects

For the initial dataset with six subjects, the 15% HoV set contained a total of 1396 samples of which 732 samples were of phase A and 664 of phase B. To overcome this unbalanced hold-out, the model was also subjected to 5-fold and 10-fold stratified CV (as mentioned earlier in section 2.2). The test set for each fold in 5-fold CV contained 1862 samples, with 931 samples of A and B phases each whereas each fold in 10-fold CV contained 931 samples, with 465 samples of A and 466 of B phases.

Method	Training Accuracy	Validation Accuracy	Testing Accuracy
15% HoV	74.02%	72.05%	**74.64%**
5-Fold CV	$76.07 \pm 0.29\%$	$73.15 \pm 0.58\%$	$\mathbf{72.27 \pm 0.69\%}$
10-Fold CV	$76.04 \pm 0.84\%$	$72.66 \pm 0.25\%$	$\mathbf{72.66 \pm 0.10\%}$

Table 5: Accuracies recorded for various validation methods performed on initial dataset.

The training, validation and testing accuracies for all the validation methods applied on the initial dataset have been tabulated in Table 5. The model gave the highest training accuracy of 74.02%, validation accuracy of 72.05% and scored 74.64% in 15% HoV method. An average training accuracy of 76.07 ± 0.29% and an average validation accuracy of 73.15 ± 0.58% was recorded for 5-fold CV, resulting in an average testing accuracy of 72.27 ± 0.69%. The results for 10-fold cross CV were nearly equivalent to that of 5-fold, giving an average training accuracy of 76.04% ± 0.4%, an average validation accuracy of 72.66% ± 0.25%, and resulting an average score of 72.66% ± 0.10%.

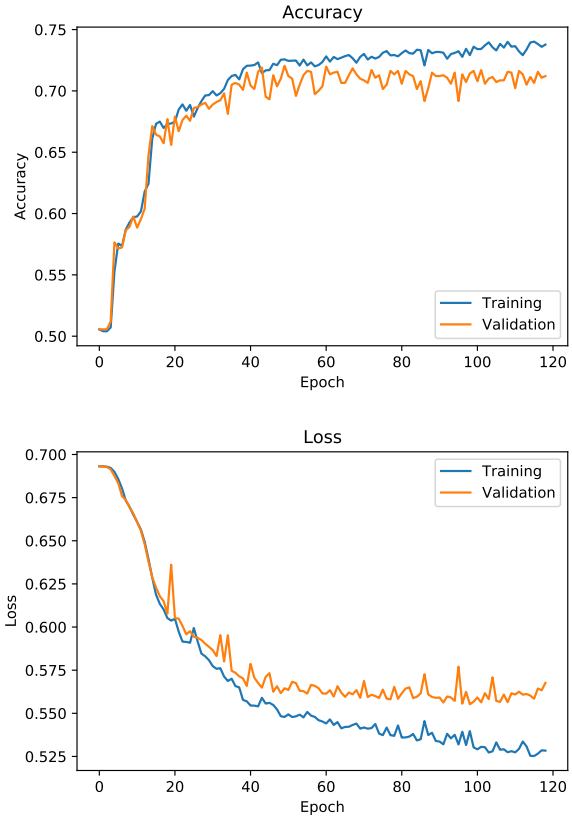

Figure 6: Accuracy and Loss graphs for 6 subjects.

The plots for the training and validation accuracy versus epochs and loss versus epochs for HoV method are shown in Fig. 6. No under-fitting or overfitting can be observed from the graphs. The training process was early stopped at 119th epoch. This means the model reaches minimum validation loss at the 99th epoch and the evaluation process was done using the weights at this stage.

Classification capabilities of an algorithm can be well interpreted from confusion matrix. Performance parameters such as sensitivity, false positive/negative rate, etc., can be derived from the confusion matrix. The confusion matrix for 15% HoV is shown in Fig. 7. As observed, 43.62% of the samples were predicted correctly as A phase and 31.02% of the samples as B phase.

Other performance parameters for all the validation methods have been summarized in Table 6. It can be noted that the model is highly capable of classifying A phases with a sensitivity of 83.20% in HoV method and 82.23 ± 1.36% in 5-fold CV although 10-fold CV gave an unexpected lower sensitivity of 78.42 ± 2.56%. In contrast, the model's capability for classifying B phases is fairly moderate. The HoV method resulted in a specificity of 65.21% while 5-fold and 10-fold CV resulted in 62.32 ± 1.28% and 66.9 ± 2.22% respectively.

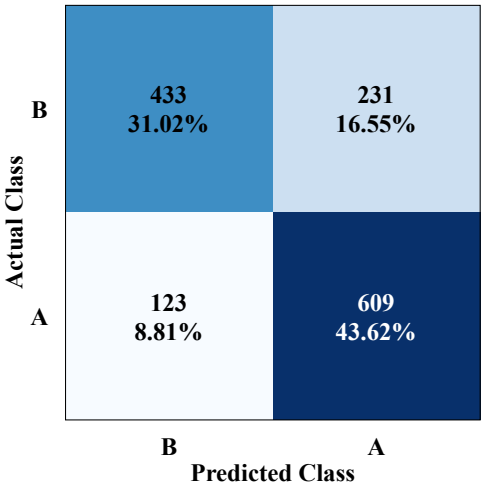

Figure 7: Confusion matrix for 15% HoV method for 6 subjects.

Method	Precision	Sensitivity	Specificity	F1 Score
15% HoV	72.50%	83.20%	65.21%	77.48%
5-Fold CV	68.54 ± 0.66%	82.23 ± 1.36%	62.32 ± 1.28%	74.82 ± 0.69%
10-Fold CV	70.34 ± 1.06%	78.42 ± 2.56%	66.9 ± 2.22%	74.14 ± 1.2%

Table 6: Other performance parameters recorded for initial dataset

3.2. Extended dataset with 14 subjects

The 15% hold-out of the extended dataset contained 3298 samples with 1659 samples of phase A and 1639 samples of phase B. Each fold of 5-fold CV contained 4397 samples, with 2199 samples of A and 2198 samples of B phases whereas each fold of 10-fold CV contained 2199 samples, with 1099 samples of phase A and 1100 of phase B.

The training, validation and testing accuracies for all the validation methods applied on the extended dataset have been tabulated in Table 7.

Results	Training Accuracy	Validation Accuracy	Testing Accuracy
15% HoV	69.83%	68.37%	**68.07%**
5 Fold CV	73.17 ± 1.41%	66.86 ± 0.82%	**67.64 ± 0.71%**
10 Fold CV	74.18 ± 1.42%	68.41 ± 0.39%	**67.76 ± 0.63%**

Table 7: Accuracies recorded for various validation methods performed on extended dataset.

The highest training and validation accuracy recorded for HoV method were 69.83% and 68.37% respectively and re-

sulted with a testing accuracy of 68.07%. Following up with 5-fold CV, we recorded the maximum metrics of each fold which resulted in a training accuracy of 73.17 ± 1.41% and a validation accuracy of 66.86 ± 0.82% with a testing accuracy of 67.64 ± 0.71%. On the other hand, for 10-fold CV we recorded a training accuracy of 74.18 ± 1.42% and a validation accuracy of 68.41 ± 0.39% with a testing accuracy of 67.76 ± 0.63%.

The plots for HoV method applied on this extended dataset are shown in Fig. 8. It can be observed that the model was early stopped at 106th epoch which means that the minimum validation loss occured at the 86th epoch and the model was evaluated using the weights from this checkpoint.

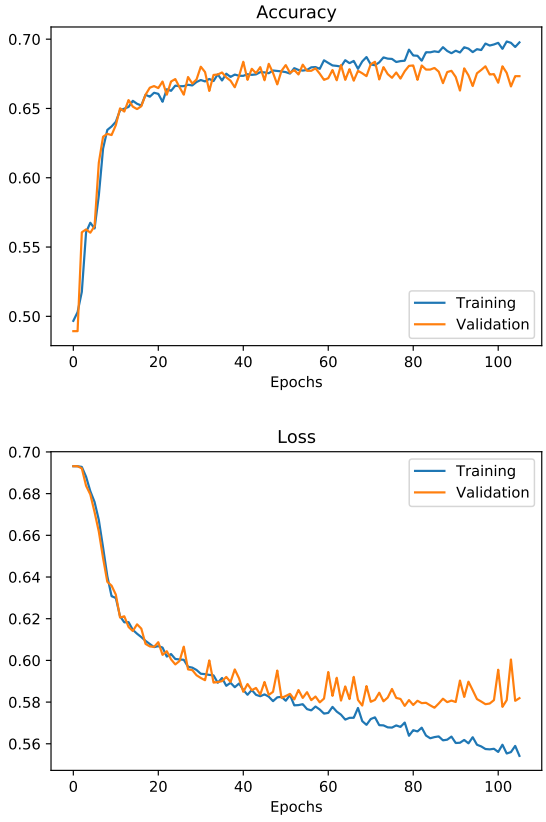

Figure 8: Accuracy and Loss graphs for extended dataset.

The confusion matrix can be for this evaluation method can be seen in Fig. 9. As observed, the model was able to correctly predict 36.96% of the A phases and 31.11% of the B phases.

Other performance parameters have been recorded in Table 8. It can be noted that the model's sensitivity to A phases albeit significantly reduced, is still strong at 73.48% for HoV method with 5-fold and 10-fold CV methods almost equivalent at 72.89 ± 3.14% and 72.25 ± 2.84% respectively. Even with the sample size increased, our model's capability of classifying B phases (specificity) still remains equivalent to that of the initial dataset, with 62.60% in HoV method and 62.41 ± 2.31% and 63.40 ± 2.78% for 5-fold and 10-fold CV respectively,

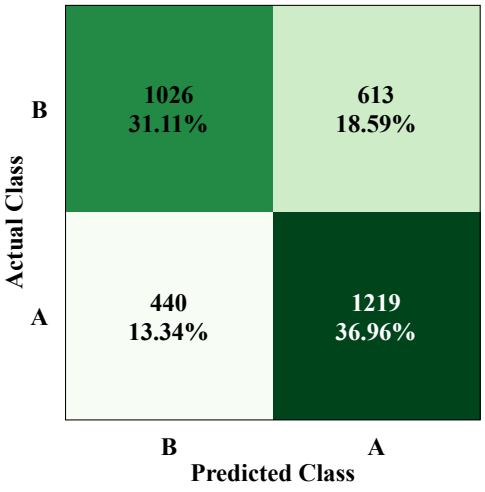

Figure 9: Confusion matrix for 15% hold out validation method for extended dataset.

Parameters	Precision	Sensitivity	Specificity	F1 Score
15% HoV	66.54%	73.48%	62.60%	69.84%
5 Fold CV	65.99 ± 0.63%	72.89 ± 3.14%	62.41 ± 2.31%	69.23 ± 1.27%
10 Fold CV	65.46 ± 2.76%	72.25 ± 2.84%	63.40 ± 2.78%	68.63 ± 1.76%

Table 8: Other performance parameters recorded for extended dataset.

The results obtained by previous researches have been summarized in Table 9 for comparison with the results presented in this work. It can be noted that we have achieved the highest accuracy and sensitivity with minimum deviation on a balanced dataset consisting of 9306 samples.

4. Conclusion

Cyclic alternating patterns are useful for the analysing the sleep of a person and can be used to identify sleep disorders such as narcolepsy, bruxism, sleep apnea etc. In this paper we present a one of its kind one-dimensional convolutional recurrent neural network that hybridizes dimensionality reducing Inception modules with gated recurrent units. This approach allowed us to significantly reduce the heavy computational cost required for feature extraction without compromising the accuracy of the model. The data set acquired was balanced with equal number of A and B phases samples to reduce bias to a particular class during training. Initially a dataset with 6 subjects was considered since their data was consistent with a sampling frequency of 512 Hz. Later the data for the remaining 8 subjects was re-sampled to match the frequency and to increase the sample size of our dataset. The model was evaluated using a 15% hold-out validation method, along with a 5-fold and 10-fold stratified cross validation method to produce results without any overfitting. We were able to achieve an accuracy of 74.64% on the initial dataset and 68.07% on the extended dataset using 15% hold-out method. The proposed model is highly capable of classifying A-phases with a sensitivity of 83.20% on

Author	Approach	Sampling frequency (Hz)	Number of samples	Performance parameters (%)		
				Accuracy	Specificity	Sensitivity
Mendez et al. [24]	K-nearest neighbours	1000	Unbalanced phase A - 3963	80.00	80.00	70.00
Mariani et al. [36]	Artificial neural network	100	Unbalanced 240429	87.19 ± 2.48	90.49 ± 2.80	69.55 ± 6.60
Navona et al. [37]	Thresholding	128	Unbalanced	77.00	90.00	80.00
Mariani et al. [25]	Support vector machines, linear discriminant analysis	100	Unbalanced	84.90 ± 4.80	86.60 ± 6.30	72.50 ± 10.90
Hartmann et al. [38]	Variable long short-term memory network	1000	Balanced & Unbalanced	82.42 ± 6.59	83.90 ± 8.95	75.28 ± 12.00
Dhok et al. [28]	Wigner-Ville based Rényi entropy features, Gaussian SVM	512	Balanced A-phase: 4653 B-phase: 4653	72.35 ± 0.20	69.19 ± 0.30	76.76 ± 0.2
			Unbalanced A-phase: 9052 B-phase: 62880	87.45 ± 0.20	52.09 ± 0.10	87.75 ± 0.20
Dhok et al. [10]	1D-CNN	512	Balanced A-phase: 4653 B-phase: 4653	73.64	66.95	80.29
This work	**Inception based 1D-CNN, gated recurrent units hybrid network**	**512**	**Balanced A-phase: 4653 B-phase: 4653**	**74.64**	**65.21**	**83.20**
			Balanced (extended) A-phase: 10992 B-phase: 10992	**68.07**	**62.60**	**73.48**

Table 9: A Summary of all the works on the EEG datasets

the initial and 73.48% on the extended dataset. Comparatively, the model's capability of classifying B phases is subpar with a specificity of 65.21% on initial and 62.60% on extended dataset.

We intend to improve and extend our work in the following ways for the foreseeable future:

- Preprocess the samples to reduce noise.
- Deploy the proposed model along with modifications for a multi-class classification of patients with sleep disorders such as insomnia, narcolepsy, bruxism, nocturnal frontal lobe epilepsy etc.
- Explore data from other channels like EOG, EMG, ECG available in [26, 27].

Furthermore the model needs to be improved more before its clinical deployment which would allow experts for an easier diagnosis.

References

[1] Simon Hartmann and Mathias Baumert. Automatic a-phase detection of cyclic alternating patterns in sleep using dynamic temporal information. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(9):1695–1703, 2019. doi: 10.1109/TNSRE.2019.2934828.

[2] CDC - How Much Sleep Do I Need? - Sleep and Sleep Disorders, 03 2017. URL https://www.cdc.gov/sleep/about_sleep/how_much_sleep.html.

[3] Paula Alhola and Päivi Polo-Kantola. Sleep deprivation: Impact on cognitive performance. *Neuropsychiatric disease and treatment*, 3 (5):553–567, 2007. ISSN 1176-6328. URL https://pubmed.ncbi.nlm.nih.gov/19300585. 19300585[pmid].

[4] Sleep statistics: Data about sleep disorders | american sleep association. https://www.sleepassociation.org/about-sleep/sleep-statistics/. (Accessed on 06/30/2021).

[5] Swapna Bhaskar, D. Hemavathy, and Shankar Prasad. Prevalence of chronic insomnia in adult patients and its correlation with medical comorbidities. *Journal of family medicine and primary care*, 5(4): 780–784, 2016. ISSN 2249-4863. doi: 10.4103/2249-4863.201153. URL https://pubmed.ncbi.nlm.nih.gov/28348990. 28348990[pmid].

[6] *Ballenger's Otorhinolaryngology: Head and Neck Surgery - John Jacob Ballenger, James Byron Snow*.

[7] Stanislas Chambon, Mathieu N. Galtier, Pierrick J. Arnal, Gilles Wainrib, and Alexandre Gramfort. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(4):758–769, April 2018. doi: 10.1109/tnsre.2018.2813138. URL https://doi.org/10.1109/tnsre.2018.2813138.

[8] Edgar R. Arce-Santana, Alfonso Alba, Martin O. Mendez, and Valdemar Arce-Guevara. A-phase classification using convolutional neural networks. *Medical & Biological Engineering & Computing*, 58(5): 1003–1014, March 2020. doi: 10.1007/s11517-020-02144-6. URL https://doi.org/10.1007/s11517-020-02144-6.

[9] A.A.O.S Medicine. The AASM manual for the scoring of sleep and associated events: Rules. *Terminology and Technical Specifications*, 2007.

[10] Hui Wen Loh, Chui Ping Ooi, Shivani G. Dhok, Manish Sharma, Ankit A. Bhurane, and U. Rajendra Acharya. Automated detection of cyclic alternating pattern and classification of sleep stages using deep neural network. *Applied Intelligence*, June 2021.

doi: 10.1007/s10489-021-02597-8. URL https://doi.org/10.1007/s10489-021-02597-8.

[11] Martin Oswaldo Mendez, Ioanna Chouvarda, Alfonso Alba, Anna Maria Bianchi, Andrea Grassi, Edgar Arce-Santana, Guilia Milioli, Mario Giovanni Terzano, and Liborio Parrino. Analysis of a-phase transitions during the cyclic alternating pattern under normal sleep. *Medical & Biological Engineering & Computing*, 54 (1):133–148, Jan 2016. doi: 10.1007/s11517-015-1349-9. URL https://doi.org/10.1007/s11517-015-1349-9.

[12] Altevogt BM Colten HR. Sleep disorders and sleep deprivation: An unmet public health problem. *Institute of Medicine (US) Committee on Sleep Medicine and Research-Washington (DC): National Academies Press (US)*, 2006. doi: https://www.ncbi.nlm.nih.gov/books/NBK19956/.

[13] Araujo JF Patel AK, Reddy V. *Physiology, Sleep Stages*. doi: https://www.ncbi.nlm.nih.gov/books/NBK526132/.

[14] Foroozan Karimzadeh, Esmaeil Seraj, Reza Boostani, and Mohammad Torabi-Nami. Presenting efficient features for automatic cap detection in sleep eeg signals. In *2015 38th International Conference on Telecommunications and Signal Processing (TSP)*, pages 448–452, 2015. doi: 10.1109/TSP.2015.7296302.

[15] Liborio Parrino, Andrea Grassi, and Giulia Milioli. Cyclic alternating pattern in polysomnography. *Current Opinion in Pulmonary Medicine*, 20(6):533–541, 2014. doi: 10.1097/mcp.0000000000000100.

[16] Anilkumar AC. Nayak CS. *EEG Normal Sleep*. doi: https://www.ncbi.nlm.nih.gov/books/NBK537023/.

[17] Reza Boostani, Foroozan Karimzadeh, and Mohammad Nami. A comparative review on sleep stage classification methods in patients and healthy individuals. *Computer Methods and Programs in Biomedicine*, 140:77–91, 2017. ISSN 0169-2607. doi: https://doi.org/10.1016/j.cmpb.2016.12.004. URL https://www.sciencedirect.com/science/article/pii/S0169260716308276.

[18] Mashael Aldayel, Mourad Ykhlef, and Abeer Alnafjan. Deep learning for eeg-based preference classification in neuromarketing. *Applied Sciences*, 10:1525, 02 2020. doi: 10.3390/app10041525.

[19] Chunxiao Han, Yaru Yang, Xiaozhou Sun, and Yingmei Qin. Complexity analysis of eeg signals for fatigue driving based on sample entropy. In *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–9, 2018. doi: 10.1109/CISP-BMEI.2018.8633011.

[20] Vijaya Kumar Gurrala, Padmasai Yarlagadda, PadmaRaju Koppireddi, and V Hari Praneet Sreenivasula. A review on analysis of sleep eeg signals. In *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pages 289–294, 2020. doi: 10.1109/ICECA49313.2020.9297564.

[21] Hui Wen Loh, Chui Ping Ooi, Jahmunah Vicnesh, Shu Lih Oh, Oliver Faust, Arkadiusz Gertych, and U. Rajendra Acharya. Automated detection of sleep stages using deep learning techniques: A systematic review of the last decade (2010–2020). *Applied Sciences*, 10(24):8963, December 2020. doi: 10.3390/app10248963. URL https://doi.org/10.3390/app10248963.

[22] Division of Sleep Medicine. Sleep division. URL https://med.stanford.edu/sleepdivision.html.

[23] European sleep research laboratories. URL https://esrs.eu/european-sleep-research-laboratories/.

[24] Martin Oswaldo Mendez, Ioanna Maria Chouvarda, Alfonso Giovanni Alba, Anna undefined Bianchi, Andrea undefined Grassi, Edgar undefined Arce-Santana, Guilia undefined Milioli, Mario undefined Terzano, and Liborio undefined Parrino. Analysis of a-phase transitions during the cyclic alternating pattern under normal sleep. *Medical Biological Engineering Computing*, 54(1):133–148, 2015. doi: 10.1007/s11517-015-1349-9.

[25] Sara Mariani, Elena Manfredini, Valentina Rosso, Andrea Grassi, Martin O. Mendez, Alfonso Alba, Matteo Matteucci, Liborio Parrino, Mario G. Terzano, Sergio Cerutti, and Anna M. Bianchi. Efficient automatic classifiers for the detection of a phases of the cyclic alternating pattern in sleep. *Medical & Biological Engineering & Computing*, 50 (4):359–372, March 2012. doi: 10.1007/s11517-012-0881-0. URL https://doi.org/10.1007/s11517-012-0881-0.

[26] Mario Giovanni Terzano, Liborio Parrino, Adriano Sherieri, Ronald Chervin, Sudhansu Chokroverty, Christian Guilleminault, Max Hirshkowitz, Mark Mahowald, Harvey Moldofsky, Agostino Rosa, and et al. Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (cap) in human sleep. *Sleep Medicine*, 2(6):537–553, 2001. doi: 10.1016/s1389-9457(01)00149-6.

[27] Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, H. Eugene Stanley, and et al. Physiobank, physiotoolkit, and physionet. *Circulation*, 101(23), 2000. doi: 10.1161/01.cir.101.23.e215.

[28] Shivani Dhok, Varad Pimpalkhute, Ambarish Chandurkar, Ankit Bhurane, Manish Sharma, and U Rajendra Acharya. Automated phase classification in cyclic alternating patterns in sleep stages using wigner-ville distribution based features. *Computers in Biology and Medicine*, 03 2020. doi: 10.1016/j.compbiomed.2020.103691.

[29] Terzano MG, Parrino L, Sherieri A, Chervin R, Chokroverty S S, Guilleminault C C, Hirshkowitz M, Mahowald M, Moldofsky H, Rosa A, and et al. Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (cap) in human sleep, 2001. URL https://pubmed.ncbi.nlm.nih.gov/14592270/.

[30] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015. doi: 10.1109/CVPR.2015.7298594.

[31] D. Rumelhart, Geoffrey E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.

[32] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Fethi Bougares, Holger Schwenk, and Y. Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. 06 2014. doi: 10.3115/v1/D14-1179.

[33] Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning, December 2014*, 2014.

[34] Keras Team. Keras documentation: Gru layer. URL https://keras.io/api/layers/recurrent_layers/gru/.

[35] Ibrahem Kandel and Mauro Castelli. The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT Express*, 6(4):312–315, 2020. ISSN 2405-9595. doi: https://doi.org/10.1016/j.icte.2020.04.010. URL https://www.sciencedirect.com/science/article/pii/S2405959519303455.

[36] Sara Mariani, Andrea Grassi, Martin O. Mendez, Giulia Milioli, Liborio Parrino, Mario G. Terzano, and Anna M. Bianchi. Eeg segmentation for improving automatic cap detection. *Clinical Neurophysiology*, 124(9):1815–1823, 2013. ISSN 1388-2457. doi: https://doi.org/10.1016/j.clinph.2013.04.005. URL https://www.sciencedirect.com/science/article/pii/S1388245713002654.

[37] Carlo Navona, Umberto Barcaro, Enrica Bonanni, Fabio Di Martino, Michelangelo Maestri, and Luigi Murri. An automatic method for the recognition and classification of the a-phases of the cyclic alternating pattern. *Clinical Neurophysiology*, 113(11):1826–1831, 2002. ISSN 1388-2457. doi: https://doi.org/10.1016/S1388-2457(02)00284-5. URL https://www.sciencedirect.com/science/article/pii/S1388245702002845.

[38] Simon Hartmann and Mathias Baumert. Automatic A-Phase Detection of Cyclic Alternating Patterns in Sleep Using Dynamic Temporal Information. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(9):1695–1703, 2019. doi: 10.1109/TNSRE.2019.2934828.