Sleep arousal detection for monitoring of sleep disorders using one-dimensional convolutional neural network-based U-Net and bio-signals

Priya Mishra and Aleena Swetapadma
School of Computer Engineering, Kalinga Institute of Industrial Technology
Deemed to be University, Bhubaneswar, India

Abstract
Purpose – Sleep arousal detection is an important factor to monitor the sleep disorder.
Design/methodology/approach – Thus, a unique nth layer one-dimensional (1D) convolutional neural network-based U-Net model for automatic sleep arousal identification has been proposed.
Findings – The proposed method has achieved area under the precision–recall curve performance score of 0.498 and area under the receiver operating characteristics performance score of 0.946.
Originality/value – No other researchers have suggested U-Net-based detection of sleep arousal.
Research limitations/implications – From the experimental results, it has been found that U-Net performs better accuracy as compared to the state-of-the-art methods.
Practical implications – Sleep arousal detection is an important factor to monitor the sleep disorder. Objective of the work is to detect the sleep arousal using different physiological channels of human body.
Social implications – It will help in improving mental health by monitoring a person’s sleep.
Keywords Sleep disorder, CNN, EEG, ECG, U-Net
Paper type Research paper

1. Introduction
Disorders caused due to insufficient sleep is a most important emerging health problem in young population in both developed economies like the USA and growing countries like India. Major causes of insufficient sleep are various disturbances which cause fragmentation of sleep due that can be identified through sleep arousals. Sleep disorders if left untreated could cause adverse domino health effects. Polysomnographic (PSG) recordings have been often used to detect various sleep arousals which are then used to monitor sleep disorders. The physiological measurements included in often used PSG recordings are electrocardiogram (ECG), electroencephalography (EEG), electromyography (EMG), oxygen saturation (SaO2) and airflow (Ghassemi et al., 2018). American Academy of Sleep Medicines has declared the rules for classifications of different types of sleep arousals like vocalizations, respiratory effort-related arousals (RERAs), leg movements, hypopneas or obstructive sleep apneas, snoring, etc. (Bonnet et al., 1992). Different methods have been suggested for sleep-related studies as discussed below.

Conflict of interest: Authors do not have any conflict of interest.
Availability of data and material: Data have been taken from database which is freely available.
Authors' contributions: P.M. – simulation, manuscript writing
A.S. – concept, manuscript writing, supervision
Cardiorespiratory signals such as heart rate, respiratory effort and ECG for classification of sleep and awake using recurrence quantification analysis have been used in Rolink et al. (2019). According to this method, the hypothesis classification of sleep and wake is more difficult and shows less performance in elder group than younger group. Discriminant atom selection–k-means singular value decomposition (DAS-KSVD) for classifying normal breathing and abnormal respiratory events in multi-class using SaO2 signals has been suggested in Rolon et al. (2020). The main objective was to compare the performances of classification with most discriminative columns selection, most discriminative atoms selection and DAS-KSVD using obstructive sleep apnea–hypopnea syndrome screening. Ensemble empirical mode decomposition (EEMD) and complete EEMD with adaptive noise (CEEMDAN) algorithms were used for extraction of respiration signal from an ECG in Zarei and Asl (2020). In Swetapadma (2017), a method has been proposed for sleep disorder monitoring using a finite-state machine taking EEG signals as input.

Different machine learning algorithms deployed for classification of RERA and non-RERA include support vector machine (SVM), k-nearest neighbor, Gaussian SVM, decision tree and used bagging, boosting methods like AdaBoost to enhance the accuracy (Karimi and Asl, 2021). In Bhattcharjee et al. (2018), an unbiased random forest feature extraction technique for multiple physiological signal has been proposed that mainly extracts generic and domain-specific features. In Lazazzera et al. (2020), various machine learning classifiers were used for classifying hypopnea and sleep apnea using peripheral SaO2 and photoplethysmography has been proposed. In Swetapadma and Swain (2016), a comparative study of various data mining method has been carried out for sleep wave and sleep stage classification.

Some appropriate approaches for classification of the arousals are made through convolutional neural network (CNN). The scattering transform of raw signals has been suggested in Warrick and Homsi (2018) which applied three layers of long short-term memory (LSTM) for sequence learning. In Howe-Patterson et al. (2018), architecture consisting of multiple dense convolutional units with bidirectional LSTM layers for detection of sleep arousal using PSG has been suggested. It excluded ECG signals and used the remaining 12 PSG signals. The methodology proposed in Jia et al. (2018) is data-driven and EEG-oriented method. The only EEG raw signals were partitioned into different segments of 10 s and power spectral density (PSD) was also computed for each segment. For arousal classification, Each PSD is fed to convolution neural network. After applying dropout layer and regularization, it was unable to solve the distribution difference problem. An ensemble CNN known as Siamese Net for detection of multimodal signal has been proposed in Erdamar and Aksahin (2020). The various raw signals like EEG, EMG, airflow and SaO2 are directly applied to neural network. In related studies, some other methods are deployed by other researchers. In Zhou et al. (2020), convolutional-residual network with positional embedding and multi-head attention for automatic sleep arousal detection has been proposed. The main focus was to extract features single channel one by one. A bidirectional gated recurrent unit (GRU) network architecture deployed in Casal et al. (2021) for classifying the sleep stages in asleep or awake. Excluding EEG signals, heart rate and SaO2 are taken as inputs and fed the signals into two stacked layers of bidirectional GRU.

The method proposed in Liu et al. (2020) for detection of sleep arousals implemented multiple convolutional network and random forest using 8 representative signals from 13 channels of PSG. It concludes detecting sleep arousals for few patients is not appropriate. ResNet50 combined with preprocessing like data augmentation for distinction between corona virus pneumonia and other pneumonia has been suggested in Zheng et al. (2021). In Jiang et al. (2020), different deep learning methods for image segmentation detect the
localization of white matter hyper intensities on magnetic resonance imaging images. In Hua et al. (2019), CNN and brain connection-based neural network are proposed for detection of proficiency of the subject using EEG signals. Dreem one shot event detector along with deep learning has been proposed in Chambon et al. (2019) to identify the sleep disorder.

In Shen (2018), a method using CNN used normalization, pooling, activation and dropout techniques. It has area under the receiver operating characteristic (AUROC) score of 0.514293 ± 0.054509 and area under the precision–recall curve (AUPRC) performance score of 0.501947 ± 0.063199. In He et al. (2018), a method was developed to detect non-apnea sources of arousals. A deep learning method that consists convolutional layers with residual connections, an LSTM layer and two fully connected layers has been used. In Li et al. (2018), a 35-layer CNN with a linear spatial filtering with 1 CNN layer, 33-layer residual networks and 1 fully connected layer has been proposed for sleep arousal detection. In Schellenberger et al. (2018), LSTM has been used for automatic arousal detection. The AUPRC score was 0.14 for the LSTM-based method. In Prainsson et al. (2018), a method for classifying target sleep arousal regions has been proposed. Time and frequency domain features are fed into a bidirectional recurrent neural network using LSTM nits. In Li and Guan (2021), a deep CNN has been proposed for detection of sleep arousal. It automatically segments sleep arousal regions in a sleep record based on the corresponding PSG signals.

In the above described methods, most of the researchers use different classifier for detection of sleep arousal regions. Plenty of researches about signal processing and artificial intelligence (AI) focus on physiological signals, which push medical AI to a new climax. The main objective of this work is to detect the sleep arousal correctly from the available signals. The recordings were of 8 h sleep record appraised 200 Hz with 13 physiological signals which is polysomnographic (PSG) recordings. There are numerous important queries with the growth in learning technologies that are well suited for arousal detection. In this study, well-known deep learning techniques have been proposed for classifying non-apnea arousals. These techniques have proven to be effective in feature extraction, which is at the core of a useful classification algorithm. The rest of the paper is structured as follows. Section 2 describes about the U-Net architecture, Section 3 outlines the proposed model architecture and experimental settings, Section 4 describes the experimental results, Section 5 includes the discussion and Section 6 concludes the paper with future work directions.

2. U-Net architecture

Deep learning techniques has been used efficiently now a days in signal processing, image processing, image classification and computer vision (Hong et al., 2020, 2020; Yao et al., 2023; Li et al., 2023). It can model the most informative features in an implicit way without monotonous crafting features. U-Net is a fully connected CNN deep learning technique. An expansive path as well as a contracting path is present in case of U-Net. For capturing the context, contracting path is used and for symmetric expands, expansive path is used. In the contracting path, each convolution as neural network contains repeated number of convolutions, rectified linear unit (ReLU) activation functions and max pooling strategy. The feature information is maximized during expansion while the spatial information gets minimized during contraction. During downsampling operation, the ReLU activation functions are followed as 3 × 3 in each layer and max-pooling operation are mainly used as pooling strategy (Navab et al., 2015). The U-Net architecture is also known as sliding window CNN. In this architecture, the context propagated to higher resolutions are allowed by the networks in the upsampling part. For this reason, the contracting path and expensive path are symmetric and the networks have in U-shaped architecture. The valid part of each convolution is accepted as there is no fully connected network in U-Net.
In downsampling, the feature channels are doubled, while in upsampling, the feature channels are halved. The feature maps are cropped from the contracting path in concatenation layer. As in every convolution layer, there is a loss of border pixels; the cropping of feature maps is required. For each 64-component, feature vectors are mapped according to the desired classes; for this purpose, $1 \times 1$ convolution is used in the final layer. The expansive and contractive paths are depicted in Figure 1. The downsampling blocks contains two operation blocks, i.e. one-dimensional (1D) convolution block and downsampling operations. The downsampling and upsampling blocks starting from 1 to $n$ are the number of layers used in U-Net. Both sampling blocks are connected by one 1D convolution, as shown in Figure 1.

3. Proposed method for sleep arousal detection

The proposed method for sleep arousal detection consists of different stages. The first stage of the proposed method is the selection of input and extraction of suitable features that help in decision-making. Next stage is the use of the obtained features for classification using the U-Net deep learning method. The proposed method has been described in detail in the following subsections.

3.1 Selection of input
In this work, Massachusetts General Hospital’s (MGH) computational clinical neurophysiology laboratory (CCNL), and the clinical data animation laboratory (CDAC) dataset have been used.
for validation of the method (Goldberger et al., 2000). It includes 994 PSG records, and each time point in the record is annotated as sleep (0), arousal (1) and non-scoring regions (−1). It used a train-test framework, in which 70 per cent of the dataset was used for training and rest 30 per cent of the dataset. During training process, the non-scoring regions which are annotated as (−1) are excluded. In each PSG record of MGH dataset, 13 physiological signals are provided sampled at 200 Hz. It includes seven channels of EEG signals (C3-M2, C4-M1, F3-M2, F4-M1, O1-M2, E1-M2 and O2-M1); three channels of EMG signals (chin, abdominal and chest movements); respiratory airflow signal; SaO₂ signal; and one ECG signal. Figure 2 shows the seven channels of EEG signals for normal and aroused conditions. In Figure 2, sample number 1 to 2,000 denotes normal condition and 2,001 to 4,000 denotes sleep arousal condition. Figure 3 shows the three channels of EMG signals, respiratory airflow signal, SaO₂ signal and ECG signal for normal and aroused conditions. In Figure 3, sample number 1 to 2,000 denotes normal condition and 2,001 to 4,000 denotes sleep arousal condition.

3.2 Design of sleep arousal detection method

U-Net is an U-shaped architecture with different layers of CNN. Initially, the original U-Net was a two-dimensional (2D) CNN, especially for 2D image segmentation (Ye and Sung, 2019). In this work, the structure from 2D to 1D CNN is transfigured for time-series sleep records. In this work, U-Net has been used for sleep arousal detection as signal classifier, as shown in Figure 4. The proposed model captures both long-range and short-range data. This paper proposed 1D CNN-based U-Net, which includes a total of 35 convolutional layers to capture the long-range information between data points over various timescales. Initially, 13 channels are fed as input of the proposed architecture. The proposed architecture

![EEG signals of seven channels](image-url)

**Source:** Figure by authors
comprises two main components – the encoder and the decoder. The encoder accepts a $2^{23}$ time points in PSG sleep record and encodes it into a latent space annotated as the red dotted rounded rectangle, as shown in Figure 4. Irrespective of the original lengths of sleep records, these are centered, within the 8-million input spaces by appending with 0s at the end. Specifically, the encoder ensures three operations, i.e. convolution–convolution–max-pooling which are used to reduce the size from $2^{23}$ to $2^8$. To encode the increased information, the number of channels is raised from 13 to 480, countervailing time-domain resolution losses. In each convolutional layer, to collect neighborhood information, the convolution operation is performed on data along the time axis. This encoded information in each layer is unique due to the different data sizes in the convolutional layers. This encoder therefore enables us to learn the interactions between data points on various timescales. Decoder, on the other hand, is used to decrypt compressed data from the latent space (He et al., 2018). Specifically, the decoder ensures three operations, i.e. convolution–convolution–upsampling which are employed to decrease the number of channels and to increase the size of the data. Furthermore, concatenation operation is performed by incorporating the information at each timescale from both the encoder and the decoder. Finally, the architecture produces a segmentation of the sleep records, with low prediction values indicating sleep and high prediction values indicating arousal.

All the experiments are implemented on system which supports Jupiter Notebook with Intel(R) Core(TM) i5-6,700 K CPU @ 4.00 GHz CPU, NVIDIA GeForce RTX GPU and 32 GB RAM. For all experiments, kernel size of 7 is used in the convolution operation because increase in the kernel size, the performance is not affected. Batch normalization is employed after each convolutional layer. In addition, after each convolution operation, rectified linear unit (ReLU), non-linear activation function is employed which allows for quick and efficient training of neural networks. In the final output layer, the sigmoid activation function is
used. The Adam optimizer is employed during the training process. After designing the proposed U-Net-based method, it is tested and validated. The results obtained in the proposed method are discussed in the next section.

4. Results
The proposed U-Net-based method has been validated, and the performance has been assessed using different features as input and parameters of U-Net. Different evaluation metrics such as precision ($p_k$), recall ($r_k$), specificity ($s_k$), accuracy ($A$), AUPRC and AUROC have been used.

\[ p_k = \frac{|T \cap P_k|}{|P_k|} \quad (1) \]

\[ r_k = \frac{|T \cap P_k|}{|T|} \quad (2) \]

\[ s_k = \frac{|N \cap S_k|}{|N|} \quad (3) \]

where $T$, $N$, $P_k$ and $S_k$ are the actual target arousal regions, actual nontarget arousal regions, predicted target arousal regions and predicted nontarget arousal regions respectively. Here ($k/1,000$) is the stride of different threshold value. The AUPRC, AUROC and accuracy are calculated as follows:

**Source**: Figure by authors

![U-Net architecture of the proposed method](image-url)
4.1 Performance varying layers

The performance of the proposed 1D U-Net for sleep arousal detection has been evaluated varying number of layers. Based on the three parameters, the network is evaluated, i.e. encoder, decoder and concatenation as shown in Table I. As shown in Table I, for encoder – five convolutions, for decoder – five convolutions and for concatenation – one convolution are taken. Similarly for 13 layers, the layers of concatenation are increased by 2. Similarly for 35 layers, for encoder – 15 convolutions, for decoder – 15 convolutions and for concatenation – 5 convolutions are taken which has highest accuracy of 93.46 per cent. Hence, this 35-layer U-Net has been chosen as the optimal 1D CNN for the validation of the proposed method.

4.2 Performance varying input features

The performance of the proposed method has been evaluated using various inputs for all 13 types of sleep signals. The accuracy of the proposed method has been given in Table II using various input signals. From Table II, it can be observed that the accuracy is highest if all the signals are taken as input. Figure 5 shows the visualization of AUPRC and AUROC using various combinations of PSG recordings. Figure 6 shows the visualization of overall accuracy using various combinations of PSG recordings. Among of these, the combination of EEG, EMG, ECG, airflow and SaO2 signals in our proposed method has proved to be the accuracy with 93.46 per cent, AUPRC of 0.498 and AUROC of 0.946.
4.3 Performance of the method in time window

The performance of the proposed method has been checked in various time windows as per real scenarios. The entire time window shows good performance when tested with the proposed method. Figure 7 shows the result of the proposed method in time window. Figure 7(a) shows a time window of 5 s for normal sleep condition with actual output and predicted output. Figure 7(b) shows a time window of 5 s for sleep arousal condition with actual output and predicted output. Figure 7(c) shows a time window of 5 s which contains 2 s of normal sleep condition and 3 s of sleep arousal condition with actual output and predicted output. It can be observed that the proposed method detects the sleep arousal correctly.

4.4 Analysis of sleep using the proposed method

The performance of the method has been tested with various patients for analyzing the sleep arousal in hours of sleep. Results shows that the proposed method find the sleep arousal percentage in the sleep hours from which quality of sleep can be analyzed. Figure 8 shows the analysis of hourly sleep of a patient using the proposed method for 6 h. From Figure 8 it can be observed that third hour of sleep has more sleep arousal among all the hours followed by second hour. Using the proposed method, the sleep arousal period from person’s total sleep can be analyzed. Based on the period of sleep arousal, it can be concluded if the person has sleep disorders or not.

<table>
<thead>
<tr>
<th>Input features used</th>
<th>Overall accuracy</th>
<th>AUPRC</th>
<th>AUROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>EEG</td>
<td>90.21</td>
<td>0.487</td>
<td>0.912</td>
</tr>
<tr>
<td>EEG, ECG</td>
<td>91.32</td>
<td>0.497</td>
<td>0.909</td>
</tr>
<tr>
<td>EEG, EMG</td>
<td>90.67</td>
<td>0.496</td>
<td>0.912</td>
</tr>
<tr>
<td>EEG, EMG, ECG</td>
<td>91.78</td>
<td>0.490</td>
<td>0.932</td>
</tr>
<tr>
<td>EEG, SaO2</td>
<td>91.89</td>
<td>0.489</td>
<td>0.941</td>
</tr>
<tr>
<td>EEG, EMG, ECG, SaO2</td>
<td>92.22</td>
<td>0.491</td>
<td>0.923</td>
</tr>
<tr>
<td>EEG, airflow, ECG</td>
<td>92.56</td>
<td>0.492</td>
<td>0.912</td>
</tr>
<tr>
<td>EEG, EMG, ECG, airflow, SaO2</td>
<td>93.64</td>
<td>0.498</td>
<td>0.946</td>
</tr>
</tbody>
</table>

Source: Table by authors

Table II. Performance varying of different input feature

Figure 5. (a) AUPRC of 35-layer U-Net (b) AUPRC of 35-layer U-Net
Figure 6. Overall accuracy using different features

Source: Figure by authors

Figure 7. Performance of the method in time window

Source: Figure by authors
5. Discussion

With the advent of various powerful deep learning algorithms and open science and computational collaboration projects, the ways by which these algorithms are applied to various problems are evolving at a rapid pace. Various contextual questions based on applications of neural networks in this field are which types of algorithms and data processing methods are well suited for arousal detection, how does the length of context influence the prediction outcome, how does the input length of PSG record influence the predictions, which types of physiological signals should be used, etc. To provide an insight into these questions, we provide Table III which describes a comparative study of various works done by different researchers over the years studied about sleep stages and sleep arousal detection. In Table III, a comparative study of various existing methods along with the proposed method has been given. Most of the studies are carried out using CNNs to enhance the sleep quality. Many researchers have used various deep learning techniques for monitoring the sleep quality. The comparative study shows about the dataset used, algorithm used and AUPRC and AUROC (Table III). In Table III, it can be observed that the proposed method gives the highest AUPRC for sleep arousal detection. The importance of the proposed method can be outlined as follows:

- CNN used in this work is particularly effective at capturing hierarchical features from complex data. In the context of sleep arousal detection, it can learn to extract relevant features, which are crucial for identifying patterns associated with sleep arousals.

Figure 8. Analysis of sleep using the proposed method

Source: Figure by authors
• U-Net used in this work is designed to handle and integrate this multimodal data efficiently.

• U-Net architecture used in this work helps in segmenting EEG or other physiological signal data to identify specific regions or patterns associated with sleep arousals.

• U-Net model is optimized for real-time or near-real-time processing, making them suitable for continuous sleep monitoring.

• Each layer in the U-Net architecture plays a specific role, contributing to the overall performance of the network.

6. Conclusion
In this work, CNN layers of 1D in U-Net for sleep arousal detection using combinations of various PSG signals have been proposed. The results are clearly demonstrating the improvement of accuracy in proposed sleep arousal detection over existing methods. In this analysis, the proposed method has been tested the performance through various CNN layers of encoder, decoder and concatenation with all 13 signals among which 35 CNNs have achieved the highest accuracy of 93 per cent, AUPRC of 0.498 and AUROC of 0.946. The proposed method can be used effectively for analysis of sleep to find if the person has sleep disorders or not. Future scope of the work is to find appropriate features for sleep arousal detection using feature extraction techniques.

ORCID iDs
Priya Mishra ©https://orcid.org/0000-0002-1078-3294
Aleena Swetapadma ©https://orcid.org/0000-0001-8270-2927

References


**Method for sleep arousal detection**