ISSN: 2073-9524
Pages:113-126

Hybrid Deep Learning Model Based on Transformer Encoder for Sleep Stages Classification

Omer Abd AL-Sattar Mohammed AL-AKKAM*
Faculty of Engineering (software), Islamic Azad University of Isfahan Branch (Khorasgan), Isfahan, Iran
omerabdulstaralakam@gmail.com

Abstract

Sleep is the cornerstone of overall health, and the process of sleep staging involves classifying sleep data into specific stages. Key signals such as EEG, EOG, and EMG are useful in analysing and categorizing sleep data, but it is a complex and time-consuming task. This paper focuses on designing a hybrid deep learning model to accurately classify sleep data using the Sleep Heart Study (SHHS) dataset. Considering that sleep signals show similar temporal patterns to the time series data, we also use transform encoders to extract essential features and facilitate the discrimination of sleep stages. By leveraging the power of transform encoders to capture crucial temporal patterns, we have successfully enhanced the classification of sleep data into five stages. Through comprehensive evaluation using various criteria, we measure the performance of our model and compare it with cutting-edge methods. The results show a significant accuracy of 0.883 and 0.836 for accuracy and Cohen's kappa, respectively, confirming the effectiveness of our approach. The results also highlight the robustness and efficiency of our approach in accurately diagnosing sleep states, ultimately contributing to the advancement of sleep analysis and overall health monitoring.

Keywords: Deep learning, Transformer encoder, Signal processing, Feature extraction, Sleep classification;

***Article history: Received:** 24 June 2024; **Accepted:** 28 Jul 2024; **Published: 15 Mar** 2025*

1. Introduction

Sleep is an integral process crucial for maintaining overall health, bolstering both the body's immunity and cognitive functions. However, this essential function can be disrupted by a variety of sleep disorders, leading to diminished sleep quality and increased susceptibility to various ailments such as stroke, diabetes, and obesity [1]. Polysomnography serves as a widely employed sleep test aimed at detecting different sleep disorders [2]. Through this test, pertinent biological signals are recorded during sleep, enabling specialists to identify and address any underlying sleep-related issues. Essential signals involve an electroencephalogram (EEG), electrooculogram (EOG), electrocardiogram (ECG), and electromyogram (EMG), which are very useful in analyzing sleep data. Some common sleep disorders that can be diagnosed through polysomnography include sleep apnea, insomnia, narcolepsy, and restless legs syndrome.

By analyzing the data obtained from these biological signals, healthcare professionals can provide targeted treatments and interventions to improve an individual's sleep quality and overall well-being. It is crucial to address any sleep disorders

[*] Corresponding Author: omerabdulstaralakam@gmail.com

promptly to prevent further health complications and ensure optimal functioning during waking hours.

Sleep staging, a pivotal aspect of sleep analysis, involves the classification of sleep data into distinct stages, including Wake, N1, N2, N3, and REM, by the guidelines set forth by the American Academy of Sleep Medicine (AASM) [3]. Given the complexity and time-intensive nature of this analysis, recent efforts have turned to the use of machine learning algorithms, particularly deep learning methods, to effectively characterize sleep data [4], [5], [6], and [7].

They author in recent years, deep learning models have shown promising results in automating the process of sleep staging, offering a more efficient and potentially more accurate alternative to manual scoring. By leveraging the power of neural networks and large datasets, researchers aim to enhance the understanding of different patterns and disorders, ultimately improving diagnostic capabilities and treatment outcomes [8], [9].

Deep learning models in the task of sleep staging have various approaches such as Recurrent Neural Networks (RNNs) [10], Convolutional Neural Networks (CNNs) [11], attention Mechanisms [12], transformer-based models [13], and hybrid architectures [14]. Each of these methods has different advantages that help to better diagnose sleep states. The attention mechanism by adding different weights to different parts of input can help the model focus on important features for accurate sleep stage classification [15].

This mechanism allows the model to dynamically weigh the importance of different temporal components in the input data, enhancing the model's ability to capture intricate patterns in sleep stage transitions. Combining attention mechanisms with RNN and CNN methods can lead to even more powerful models that excel in understanding complex temporal relationships and patterns in sleep data. By integrating these techniques, researchers aim to create more accurate and robust sleep staging systems that can provide valuable insights into an individual's sleep quality and patterns. Transformer-based models have demonstrated considerable potential in learning long-range dependencies within time series data, thereby effectively capturing the inherent sequential nature of such data [16].

Through the utilization of self-attention mechanisms, these models adeptly process input sequences, extracting pertinent information essential for sequential data analysis. Additionally, transformer architectures exhibit efficiency in processing input sequences and extracting crucial information pertaining to various sleep stages through self-attentive mechanisms[17]. By employing transformer attention mechanisms, the model can dynamically focus on different segments of the input sequence, facilitating the identification of complex relationships and patterns crucial for accurate sleep stage classification. Notably, transformer computations present an appealing option for large-scale sleep applications due to their potential for expedited training times and enhanced scalability, owing to their parallel processing capabilities [18]. The integration of transformer-based models, alongside other deep learning methodologies, holds significant promise in advancing the frontier of sleep staging technology as the field continues to evolve.

Hybrid deep learning architectures by combining different methods have some potential benefits including improved model performance, increased robustness, and the ability to leverage the strengths of multiple techniques for more accurate predictions [14]. These architectures can also provide a more flexible framework for adapting to different types of data and tasks, making them a valuable tool for a wide range of applications in fields such as natural language processing, healthcare, and computer vision [19], [20], and [21].

By integrating diverse deep learning methods under a unified framework, scientists may develop models that exhibit increased capacity to handle complex data structures and nuances. More accurate and reliable predictions may result from this, especially in situations where conventional machine-learning techniques are not effective enough. Hybrid architectures are particularly suitable for dynamic and evolving datasets because they have the capacity to adjust and change over time. Given the circumstances, the integration of many methods into a single model can create new opportunities to solve difficult problems and expand the frontiers of artificial intelligence.

According to the mentioned benefits of different methods, this paper introduces a hybrid transformer-based deep learning model using several key signals to classify sleep stages as a complex task. Our proposed methodology comprises the following steps:

1. Preparing sleep data involves preprocessing and normalization procedures.
2. Constructing CNN-BiLSTM layers integrated with an attention mechanism tailored to accommodate diverse biological signals characterized by varying sampling rates.
3. Developing a transformer encoder with optimized layers and parameters.
4. Crafting a hybrid deep learning model by integrating the transformer encoder with the CNN-BiLSTM attention architecture, tailored specifically for the classification of sleep stages.

By combining various methods, our proposed hybrid deep learning architecture offers a comprehensive approach by leveraging the strengths of each component, leading to improved performance in sleep staging tasks.

Additionally, we employ various evaluation metrics including accuracy, Cohen's kappa, recall, and F-score to assess the effectiveness of the proposed hybrid model. Furthermore, we performed a comparative analysis between the developed model and other cutting-edge techniques.

The remainder of this paper is structured as follows: Section 2 presents an overview of related works, while Section 3 details the materials and methods, encompassing the dataset utilized, the methodologies employed, and the proposed hybrid deep learning model. Section 4 presents the obtained results, followed by conclusions and future directions in Sections 5 and 6.

2. Related work

This section provides a summary of cutting-edge works based on deep learning models in the field of sleep stage classification. The study of Pei et al. [22] is focused on the importance of sleep stage classification for diagnosing and treating sleep-related diseases. It introduces a deep learning-based approach using multiple biological signals like EEG, ECG, EMG, and EOG for identifying sleep stages. Unlike previous studies that rely on manual feature extraction methods, this work combined CNNs and GRUs to automatically extract features and learn transition rules for efficient sleep stage classification.

The work of Liu et al. [23], is focused on addressing the challenges in automatic sleep staging to assist in diagnosing psychiatric and neurological disorders related to somnipathy. Researchers aim to improve the classification of sleep stages using deep learning methods. The proposed deep neural network combines MSE-based U-structure and CBAM to extract multi-scale salient waves from EEG signals, capturing transition rules between sleep stages. A class adaptive weight cross-entropy loss function is introduced to tackle the class imbalance issue in the dataset. Experimental results on three public datasets demonstrate that the model outperforms existing methods significantly, suggesting its potential to replace human experts in sleep staging tasks.

The study of Zhao et al. [24], focused on enhancing sleep staging using single-channel EEG signals by incorporating long-term temporal context information alongside short-term context. The Sleep Context Net employs a combination of CNN layers for feature extraction from each sleep stage and RNN layers to capture long-term and short-term context information in chronological order. By integrating both short and long-term contexts, the model aims to improve the understanding of sleep stage transitions within a sleep cycle. Additionally, a data augmentation algorithm is designed to preserve long-term context information in EEG signals without altering the sample count.

The significance of precisely identifying arousal and sleep phases in the diagnosis of sleep disorders, which have a substantial influence on both physical and mental health, was highlighted [25]. To tackle this issue, the study presented FullSleepNet, an innovative multi-task learning methodology that makes use of fully convolutional neural networks. FullSleepNet creates segmentation masks for arousal and sleep stage labels by processing single-channel EEG data collected over the whole night. A convolutional module for extracting local features, a recurrent module for collecting long-range relationships, an attention mechanism for concentrating on pertinent input regions, and a segmentation module for generating final predictions are the four main modules that the model includes.

The study of Kong et al. [26], introduced a novel Neural Architecture Search (NAS) framework for EEG-based sleep stage classification to address the time-consuming and laborious process of designing automatic staging neural networks by human experts. The developed NAS architecture utilizes bi-level optimization approximation for architectural search, optimizing the model through search space approximation and regularization with shared parameters among cells. The results suggest that the NAS algorithm could serve as a valuable reference for future automatic network design in sleep classification.

The work of Li, et al. [27], introduced addressed the challenges in sleep stage classification by proposing a method using EEG spectrogram data. The EEGSNet model, based on CNNs and Bi-LSTMs, extracts time and frequency features to classify sleep stages. By using Gaussian error linear units (GELUs) for CNN activation, the model's generalization ability is enhanced. Evaluation on four public databases demonstrates high accuracy, MF1 scores, and Kappa values across datasets. Notably, the developed method achieves improved classification performance on the N1 sleep stage compared to existing methods.

An innovative method for analyzing input feature maps using depth-wise separable multi-resolution convolutional neural networks is presented by LWSleepNet [28]. This design uses two convolutional kernels of varying sizes to efficiently collect information at different frequencies. Furthermore, by partitioning input data into patches and utilizing a multi-head attention technique to extract time-dependent information from sleep records, the temporal feature extraction module substantially improves the model's performance. LWSleepNet is novel in that it substitutes depth wise separable convolutions for conventional convolutional processes. With this change, the model's performance is maintained but its parameters and computational cost are drastically decreased. Sleep-EDF-20 and 78, two popular public datasets, were used to assess LWSleepNet's effectiveness.

3. Materials and Methods

This section begins with a description of the utilized dataset, the chosen signals, and the preprocessing steps. Subsequently, we delve into the various methodologies adopted in this study, encompassing CNN-BiLSTM-attention and transformer architectures. Following this, we offer a comprehensive overview of our proposed hybrid deep-learning model. At the end of this section, we detail the evaluation metrics employed to assess the performance of our proposed model. In the subsequent sections, we present the experimental setup and parameter tuning process for each architecture. We then discuss the results obtained from the experiments and provide a detailed analysis of the model's performance. Finally, we conclude this paper with a discussion of the implications of our findings and potential future research directions.

3.1. Dataset

The Sleep Heart Health Study (SHHS) dataset is a comprehensive repository comprising 5793 polysomnograms featuring various biological signals [29]. Provided by the National Heart Lung & Blood Institute, this dataset serves to investigate cardiovascular and sleep-related disorders. Within the SHHS dataset, key signals recorded for each patient include EEG, EMG, EOG, ECG, chest and abdomen, oxygen saturation (SaO2), positional data, and ambient light levels. Specifically, EEG signals are captured via two channels, acquired from C4-A1 and C3-A2, while EOG signals are obtained bilaterally from both left and right channels. The SHHS dataset offers a rich resource for researchers and clinicians seeking to delve into the intricate relationship between sleep patterns and cardiovascular health. By examining the array of signals provided, including EEG, EMG, EOG, ECG, and more, investigators can gain valuable insights into the physiological mechanisms underlying various sleep disorders and their impact on cardiovascular function. This wealth of data opens avenues for in-depth analyses and the development of novel diagnostic and therapeutic approaches in the realm of sleep medicine and cardiology.

3.2. Data preprocessing

Given that our focus is on sleep staging, we choose signals that are particularly suitable for this task, namely EEG, EOG, and EMG. From the SHHS data set, a group of 100 people was randomly selected. Since each patient's signals are recorded between 6 and 8 hours, these recordings are divided into 30-second intervals to facilitate sleep stage classification. Each segment is then carefully annotated according to the American Academy of Sleep Medicine (AASM) guidelines, identifying distinct stages including awake, REM, N3, N2, and N1 [3]. This study aims to develop a hybrid deep learning model that can accurately classify sleep stages based on the signals obtained from EEG, EOG, and EMG recordings. By leveraging the annotated data from the SHHS dataset and utilizing advanced signal processing techniques including segmentation, normalization, and noise removal, we seek to train a model that can automatically differentiate between different sleep stages with high precision. The goal is to create a reliable tool that can assist sleep specialists in diagnosing sleep disorders and monitoring patients' sleep patterns effectively.

3.3. CNN-BiLSTM-Attention

In sleep staging, CNNs can extract spatial features from multi-channel physiological signals like EEG, EMG, and EOG, offering vital insights into brain activity, muscle tone, and eye movements during sleep. By processing these signals through convolutional layers, CNNs can autonomously grasp hierarchical features representing diverse sleep stage characteristics. Conversely, RNNs like Long Short-Term Memory (LSTM) or bidirectional LSTM (BiLSTM) networks excel at modeling temporal dependencies in sequential data [30]. In sleep staging, RNNs adeptly capture the temporal dynamics of EEG signals over time, facilitating pattern learning and transitions between different sleep stages.

The incorporation of attention mechanisms further boosts the CNN-RNN hybrid model's efficacy, enabling focused attention on pertinent parts of input signals. These mechanisms assign weights to segments of the input sequence based on their relevance to current predictions, empowering the model to prioritize crucial features while filtering out irrelevant or noisy data. The combination of these methods completely changes the way sleep data is analyzed in the field of sleep research. This provides an opportunity to explore the intricacies of sleep stages and patterns in more detail using different designs. As a result of this combination of spatial and temporal processing skills, the field is moving toward more sophisticated diagnostic and therapeutic applications that enable the reading of sleep-related physiological data with greater accuracy and precision. Fig. 1 shows the developed CNN-BiLSTM-attention layers in this paper.

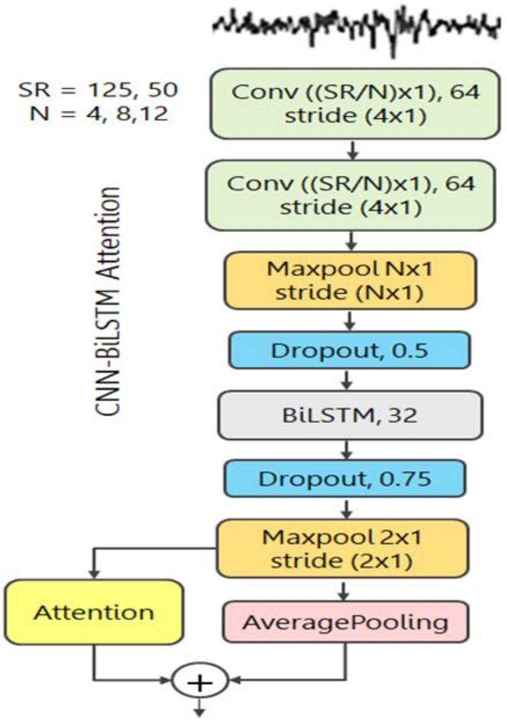

Fig. 1 Developed CNN-BiLSTM-attention layers. SR = sampling rate, and N = size of division.

The process begins with two one-dimensional convolutional layers, succeeded by max pooling and dropout operations. Subsequently, the extracted features traverse through a Bidirectional Long Short-Term Memory (BiLSTM) layer, incorporating dropout and max pooling. Finally, the output of the layers undergoes global average pooling and an attention mechanism (Dot product attention), and is combined.

This process is iterated for each signal, encompassing EEG, EOG, and EMG, each with distinct sampling rates (SR) of 125, 50, and 125, respectively. Given the variability in signal sampling rates, the kernel size and stride size are adjusted accordingly for each signal. Additionally, with N taking on values of 4, 8, and 12, three distinct features of varying sizes are extracted for each signal. For example, in EEG signals that have a sampling rate of 125, this value is divided by N to select the kernel size, resulting in 30, 15, and 10 kernels. Actually, for each signal, three high-level features based on SR and N values are extracted and then combined.

As mentioned earlier, EEG signals include two channels, C4-A1 and C3-A2, and EOG signals include left and right channels, finally, using developed layers, fifteen high-level features are extracted from EEG, EMG, and EOG signals. This comprehensive approach allows the model to effectively capture and leverage the unique characteristics of each signal type, adaptively adjusting its operations based on the specific sampling rates and feature sizes. By incorporating a diverse range of features and utilizing advanced neural network architectures such as BiLSTM and attention mechanisms, the model demonstrates robust performance in processing physiological data for sleep stage classification.

3.4. Transformer

An essential part of the Transformer architectures is the transformer encoder, which is mostly used for activities related to natural language processing but can also be modified for sequential data applications like sleep staging [13]. Transformers process long-range dependencies very well because, in contrast to recurrent or convolutional neural networks, they only use self-attention techniques to capture interactions between various items in a sequence. The design of the transformer encoder is shown in Fig. 2. It consists of several layers, such as feedforward neural networks and multi-head self-attention, followed by normalization and dropout layers.

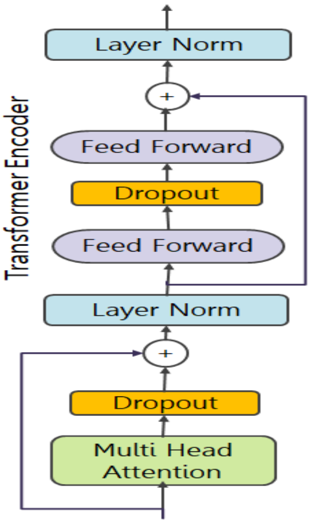

Fig. 2 The architecture of the transformer encod

This configuration enables the simultaneous processing of information by focusing each head in the transformer encoder on a specific part of the input sequence. The model can jointly attend to the data of several display subspaces at different locations thanks to the multi-head self-awareness mechanism, which improves the model's capacity to identify complex patterns and relationships in the data. This model can learn different representations and extract finer features from the input sequence by merging many heads. Because of its flexibility and adaptability, the transformer encoder is an effective tool for all kinds of sequential data, including tasks that go beyond conventional natural language processing applications.

In multi-head self-attention, the input sequence undergoes a transformation into queries, keys, and values through learnable linear projections. Attention scores are computed between queries and keys to determine key relevance, which then weights the values to generate context vectors. This process runs in parallel multiple times with distinct parameters, enabling simultaneous attention to various input segments for diverse data representations.

The final representation is usually a concatenation and linear transformation of the outputs from the multi-head self-attention process. The self-attention mechanism allows the model to preferentially focus on relevant parts of the provided data sequence while disregarding irrelevant information, resulting in an accurate categorization of sleep stages as displayed in Fig. 3.

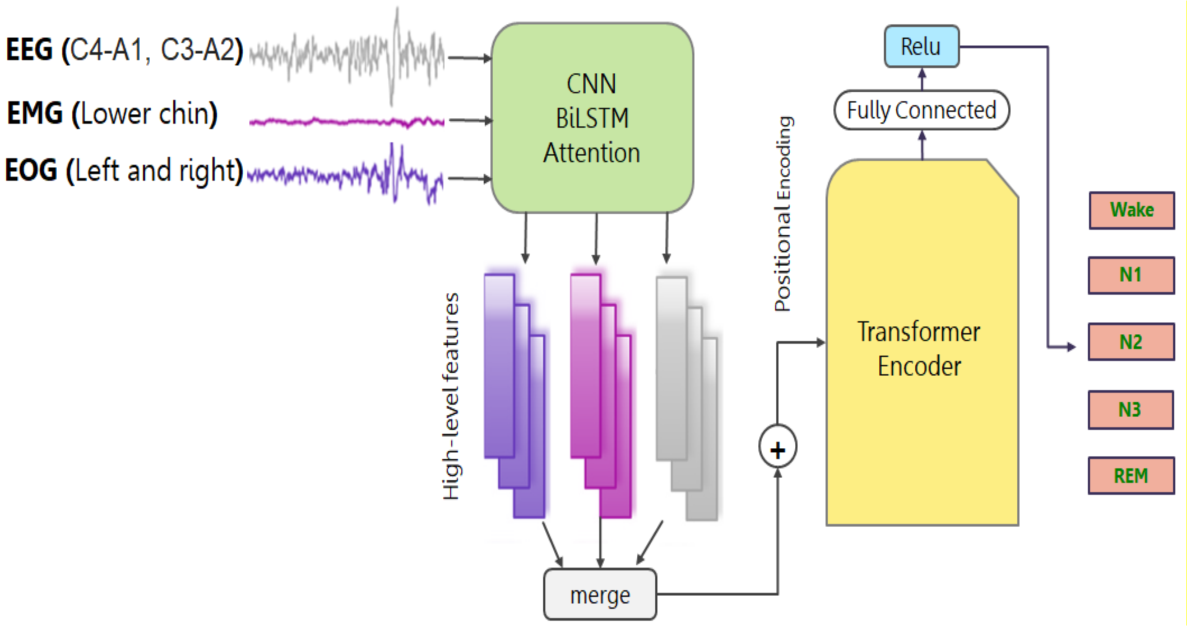

Fig. 3 The proposed hybrid deep learning model based on the transformer encoder

In the proposed model, the extracted features are fed into the transformer encoder after the combination. For the transformer architecture, we chose 4 heads of size 64 with a real activation and a dropout of 0.20. With the help of this mechanism, the model performs better in tasks such as sleep staging by better capturing complex patterns and long-range relationships in sequential data. Through the use of multi-head self-awareness, this model is able to effectively evaluate the importance of different input components and facilitate accurate understanding and flexible feature extraction. In addition, the self-attention mechanism increases interpretability by emphasizing the connections between components in the input sequence, which makes it easier to understand how the model makes decisions. For this reason, the use of self-attention has greatly improved the ability of neural networks to handle sequential inputs in different domains, demonstrating its adaptability and efficiency in challenging learning tasks.

3.5. The proposed method

This section provides a comprehensive overview of the proposed hybrid deep-learning model with an explanation of parameter settings in different parts. As mentioned earlier, we selected three key signals involving EEG (C4-A1, C3-A2), EMG (lower chin), and EOG (left and right) from the SHHS dataset. Following preprocessing and segmentation procedures, these signals are inputted into the developed CNN-BiLSTM-attention layers, as depicted in Figure 1. Within these layers, three distinct forms of high-level features are extracted from each signal, determined by the sampling rate (SR) and division size (N). The parameter settings play a crucial role in optimizing the performance of the model. In the CNN-BiLSTM-attention layers, each signal undergoes a series of transformations that enable the extraction of intricate features essential for accurate classification.

The used dot product attention helps models focus on important parts of the input by assigning different weights to different elements. It uses three main inputs: Query, Key, and Value vectors. First, it calculates the dot product between the Query vector and all Key vectors to measure their similarity. Next, it applies the softmax function to these dot product results, turning the similarity scores into a probability distribution. Finally, it uses these probabilities to weight the Value vectors and sums them up, producing a single weighted sum vector as the output of the attention mechanism. Through the integration of CNNs and RNNs with attention mechanisms, the hybrid architecture adeptly captures both spatial and temporal dependencies inherent in sleep data, culminating in enhanced accuracy in sleep stage classification.

The high-level features that have been obtained from each signal are then combined and sent to the transformer decoder part. Before that information about the positions of the tokens in the sequence is inserted using a positional encoding approach. Transformers don't naturally know the order in which tokens should appear in a series, therefore positional encoding plays a critical role in helping the model understand the token arrangement in sequence[31]. Within the context of sleep staging, positional encoding gives each token a unique representation based on where it is in the input sequence.

The positional encoding allows the transformer decoder to differentiate between tokens and learn the sequential relationships between them. This helps the model make more accurate predictions about the sleep stages based on the input signals. By incorporating positional information, the transformer can effectively capture the temporal dependencies in the data and generate meaningful output for sleep staging analysis.

Finally, the high-level features, along with their coded positions, pass through the transformer encoder and then to the fully connected layers. The Transformer encoder can handle all elements of the input sequence at once, making training much faster. It can also work with input sequences of varying lengths without needing a fixed structure.

In the end, sleep stage classification is performed using the modified linear unit (ReLU) activation function [32]. The ReLU activation function helps to introduce non-linearity into the model, allowing for more complex relationships to be captured during the classification process. This enables the model to better distinguish between different sleep stages based on the extracted features. Additionally employing the transformer encoder can effectively extract high-level features from sleep data, aiding in the discrimination of different sleep stages. The parallelizability of Transformers enables effective training on vast datasets, making them ideal for analyzing extensive sleep recordings. In summary, our hybrid deep learning model provides a robust solution for improving the accuracy of sleep staging tasks and deepening our comprehension of complex sleep stages.

3.6. Evaluation metrics

In evaluating the performance of our proposed deep learning model, we employed a range of evaluation criteria, including accuracy, Cohen's kappa, F1 score, and recall [33]. A brief overview of each criterion is provided below.

- Accuracy:

Accuracy quantifies the ratio of correctly classified instances to the total number of instances. Equation 1 represents the accuracy formula, incorporating True Positives, False Negatives, False Positives, and True Negatives as TP, FN, FP, and TN.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \qquad (1)$$

- F1 score

The F1-score represents the harmonic mean of precision and recall, offering a balanced assessment between these two metrics.

$$\text{F1-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (2)$$

- Cohen's kappa

Cohen's Kappa evaluates the agreement between true and predicted labels, considering the potential for chance agreement.

$$\text{Cohen's kappa} = \frac{p_o - p_c}{1 - p_c} \qquad (3)$$

- Recall

Recall, also called sensitivity or true positive rate, measures the proportion of true positive samples that are correctly identified by the classification model.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad (4)$$

4. Results

This section presents a series of experiments aimed at demonstrating the strengths of our proposed hybrid deep learning model. First, we assess the performance of our hybrid model in different combinations of signals. Subsequently, we present the results obtained from our developed model using multiple criteria. Finally, we perform a comparative analysis to evaluate the performance of our model against other cutting-edge methods.

4.1. Results of using different signals

The proposed model undergoes extensive training utilizing various combinations of signals to meticulously analyze their impact on sleep stage prediction. These combinations encompass (EEGs), (EEGs, and EOGs), and (EEGs, EOGs, and EMG). The detailed results of these experiments are meticulously documented in Table 1, illustrating the accuracy and F1 score achieved for each stage by every combination.

Table 1: The outcomes of the proposed model using different signals

Signals	Acc	W	N1	N2	N3	REM
EEG	0.838	0.83	0.19	0.89	0.88	0.80
EEG, EOG	0.862	0.85	0.29	0.91	0.90	0.85
EEG, EOG, EMG	**0.883**	**0.86**	**0.45**	**0.92**	**0.91**	**0.89**

As anticipated, the inclusion of additional signals correlates with a discernible enhancement in model performance. This underscores the notion that integrating multiple signals can furnish a richer and more nuanced dataset, thereby facilitating more precise and comprehensive predictions of sleep stages. Such an approach acknowledges the intricate interplay between different physiological signals during sleep, ultimately contributing to a more robust and accurate sleep staging model.

4.2. The overall results

Based on the findings presented in Table 1, we further assess the proposed model using a combination of all selected signals. The model was trained 100 epochs with a batch size of 128 and subsequently evaluated on the test dataset. Figs. 4 and 5 illustrate the corresponding learning curves and confusion matrix, respectively.

During the experiments, we randomly allocated 70% of the data to the training dataset, reserving the remaining portion for the validation and test datasets. This process is repeated ten times, and the average results were calculated for the overall assessment. Mean Square Error is utilized as the loss function, with ADAM serving as the optimizer with a learning rate of 0.0001[34], [35]. The model's hyperparameters were set to ensure convergence within a reasonable number of epochs. Regularization techniques such as L2 regularization [36] were applied to prevent overfitting, enhancing the model's generalization capabilities. Additionally, early stopping was implemented to halt training when the validation loss ceased to improve significantly, preventing potential overfitting on the training data.

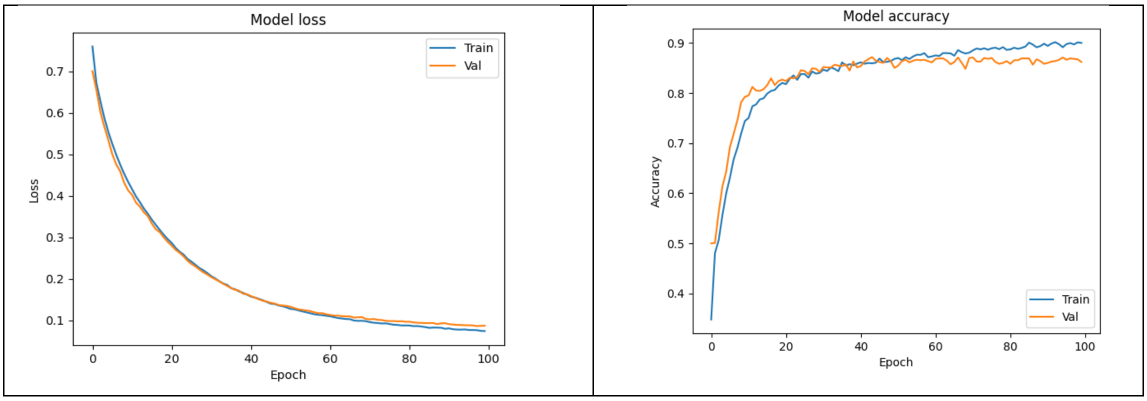

Fig. 4 Learning curves of the train and validation datasets

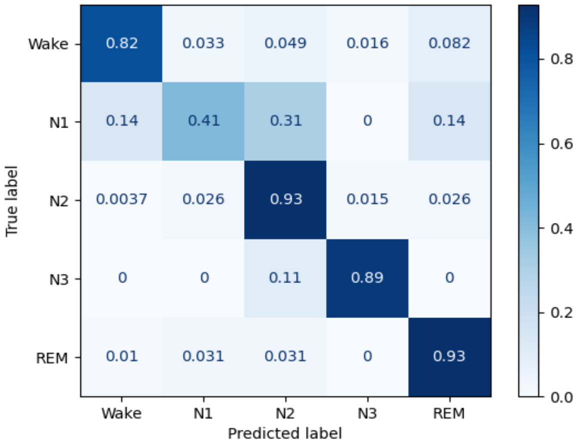

Fig. 5 Confusion matrix of the proposed deep learning model.

As depicted in Fig. 4, the accuracy and loss of both validation and training datasets are illustrated. The loss value gradually decreases to approximately 0.1, while the accuracy shows improvement, reaching 0.88. The model's performance indicates successful training with a high level of accuracy and low loss. It is worth mentioning that we have developed the proposed model step by step and analyzed its performance by adding each item. In this process, we have explored many methods and parameters that have had poor results on the performance of our model. Finally, we decided to introduce our model with the mentioned parts and settings.

In fact, by adding each part like CNN-BiLSTM layers, attention mechanism, and encoder, the results improved. CNN-BiLSTM layers, which are the basis of our architecture, improved the performance of the model to an accuracy of 0.857. By adding the attention mechanism, the accuracy reached 0.865, and finally, performance improved to 0.88 by including the encoder part. This trend suggests that the model is effectively learning from the data and making accurate predictions. Further analysis of the model's performance on unseen data will be crucial to assess its generalization capabilities.

The confusion matrix presented in Fig. 5 corresponds to the test dataset. As shown, the proposed model shows the ability to detect different stages with relatively high recall values. It is worth noting that the N1 class shows the lowest recall, which is attributed to its smaller number of samples compared to the other classes. Overall, the results indicate that the proposed model performs well in classifying the different stages, with some room for improvement in detecting the N1 class. Further analysis and augmenting data may help to enhance the models' performance in this particular class.

The model demonstrates robust performance across various metrics, with an average accuracy of 883% on the test dataset. Additionally, the model exhibits a stable training process, as depicted in Fig. 4. The confusion matrix depicted in Fig. 5, highlights the model's ability to effectively classify different signal types with high recall values. These results suggest that the proposed model, trained using a comprehensive set of signals and advanced methods, shows promising potential for real-world applications in signal processing and classification tasks.

4.3 Performance comparison

Table 2 outlines the comparative performance of our proposed deep learning model against other cutting-edge methods in sleep staging. The Table summarizes the accuracy and Cohen's kappa of different methods. All of the selected methods are based on hybrid deep learning models that are developed to classify sleep data. The table presents accuracy and Cohen's kappa values achieved by different models on the SHHS dataset. According to the table, our proposed model attained 0.883 accuracy and 0.836 Cohen kappa, demonstrating superior performance compared to other models. Cohen's kappa is used to evaluate the agreement between human ratings and deep learning models in assigning sleep stages to different epochs of sleep data.

Our deep learning model's exceptional performance on the SHHS dataset showcases its potential to significantly advance the field of sleep staging. By achieving a high accuracy of 0.883 and a substantial Cohen kappa value of 0.836, our model not only outperforms existing methods but also sets a new standard for accuracy and reliability in sleep stage classification. This success underscores the effectiveness and robustness of our proposed approach in accurately identifying and classifying sleep stages, which is crucial for improving diagnostic tools and enhancing patient care in sleep medicine.

Table 2: Performance comparison of the proposed model and other methods.

Deep Learning Models	**Methodology**	**Accuracy**	**Cohen's kappa**
Kong et al. [26]	Neural Architecture Search based on CNN	0.819	0.740
Pei et al. [22]	Combined CNNs and GRUs	0.831	0.760
Fernandez et al. [37]	Separable convolutional neural network	0.852	0.790
Liu et al. [23]	Combined multi-scale U-structure extraction and convolutional block attention module	0.868	0.835
SleepContexNet [24]	Combination of CNN and RNN layers	0.864	0.810
FullSleepNet [25]	Multi-task learning based on convolutional and recurrent modules with an attention mechanism	0.875	0.826
Pei et al. [38]	A deep convolutional neural network (CNN) combined with a long short-time memory (LSTM)	0.824	0.725
Zhang et al. [39]	A combination of an attention-based convolutional neural network and a two-branch classifier	0.881	0.835
Zhang et al. [40]	Multi-scale convolutional neural network (MSCNN) and adaptive channel feature recalibration (ACFR)	0.857	0.810
Our proposed model	Transformer-based deep learning model using CNN, BiLSTM, and attention mechanism	0.883	0.836

5. Future Directions

Incorporation of Additional Physiological Signals: Exploring the use of additional physiological signals such as heart rate variability (HRV) and respiratory patterns to further improve the accuracy and robustness of sleep stage classification. Personalized Sleep Models: Developing personalized models that take into account individual differences in sleep patterns and health conditions to provide more tailored and accurate sleep analysis.

Integration with Clinical Applications: Exploring the integration of our model with clinical applications for diagnosing and managing sleep disorders and assessing its impact on patient outcomes. Advanced Deep Learning Techniques: Continuing to explore advanced deep learning techniques, such as reinforcement learning and generative adversarial networks (GANs) to enhance model performance and address current limitations.

These additions aim to provide a clear vision for the future of our research and highlight the potential for further advancements in the field of sleep analysis and health monitoring.

6. Conclusion

This paper introduces a novel hybrid-based deep learning model for the classification of different sleep stages. We devised hybrid layers consisting of CNN, BiLSTM, and attention mechanisms to capture complex features from a combination of signals, including EEG, EOG, and EMG. After that, we used transformer encoder architecture to process these extracted features and finally predict sleep stages. We conducted extensive experiments and evaluations of the proposed model using various metrics and compared its performance with other cutting-edge methods. Our findings show that our proposed hybrid deep learning model, using advanced components and more signals, outperforms existing methods in sleep stage classification. The outcomes represent the efficiency of our model in accurately classifying sleep stages, demonstrating its potential to improve sleep monitoring and analysis in clinical and research settings. Furthermore, the power of the attentional mechanism and transformer encoder in our model provides valuable insights into the decision-making process and increases the clarity and reliability of classification results. Overall, our hybrid deep learning approach provides a promising solution to advance the field of sleep stage classification and has great potential for further exploration and development within this domain.

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

References

[1] Sateia, M.J., Buysse, D.J., Krystal, A.D., Neubauer, D.N. and Heald, J.L., "Clinical practice guideline for the pharmacologic treatment of chronic insomnia in adults: an American Academy of Sleep Medicine clinical practice guideline," Journal of clinical sleep medicine, 13(2), pp.307-349, 2017.

[2] Caples, S.M., Anderson, W.M., Calero, K., Howell, M. and Hashmi, S.D., "Use of polysomnography and home sleep apnea tests for the longitudinal management of obstructive sleep apnea in adults: an American Academy of Sleep Medicine clinical guidance statement," Journal of Clinical Sleep Medicine, 17(6), pp.1287-1293, 2021.

[3] Singh, J., Badr, M.S., Diebert, W., Epstein, L., Hwang, D., Karres, V., Khosla, S., Mims, K.N., Shamim-Uzzaman, A., Kirsch, D. and Heald, J.L., , "American Academy of Sleep Medicine (AASM) position paper for the use of telemedicine for the diagnosis and treatment of sleep disorders: an American Academy of Sleep Medicine Position Paper," Journal of Clinical Sleep Medicine, 11(10), pp.1187-1198, 2015.

[4] Mostafaei, S.H., Tanha, J., Sharafkhaneh, A., Agrawal, R. and Mostafaei, Z.H., "Biological signals for diagnosing sleep stages using machine learning models," 28th International Computer Conference, Computer Society of Iran (CSICC) (pp. 1-7). IEEE, 2023.

[5] Loh, H.W.; Ooi, C.P.; Vicnesh, J.; Oh, S.L.; Faust, O.; Gertych, A. ; Acharya, U.R.,, "Automated detection of sleep stages using deep learning techniques: A systematic review of the last decade (2010–2020)," Applied Sciences, 10(24), p.8963, 2020.

[6] Mostafaei, S.H., Tanha, J., Sharafkhaneh, A., Mostafaei, Z.H., Al-Jaf, M.H.A. and Babaei, A.F., "An Ensemble Model for Sleep Stages Classification," 31st International Conference on Electrical Engineering (ICEE) (pp. 327-332). IEEE, 2023.

[7] Mostafaei, S. H.; Tanha, J.; Sharafkhaneh, A., "A novel deep learning model based on transformer and cross-modality attention for classification of sleep stages," Journal of Biomedical Informatics, p.104689., 2024.

[8] Tsuneki, M., "Deep learning models in medical image analysis," Journal of Oral Biosciences, 64(3), pp.312-320, 2022.

[9] Mijwil, M.M.; Doshi, R.; Hiran, K.K.; Unogwu, O.J.; Bala, I.,, "MobileNetV1-based deep learning model for accurate brain tumor classification," Mesopotamian Journal of Computer Science, 2023, pp.29-38, 2023.

[10] Caterini, A.L., Chang, D.E., Caterini, A.L. and Chang, D.E.,, "Recurrent neural networks," Deep neural networks in a mathematical framework, pp.59-79, 2018.

[11] Li, Z., Liu, F., Yang, W., Peng, S. and Zhou, J., "A survey of convolutional neural networks: analysis, applications, and prospects," IEEE transactions on neural networks and learning systems, 33(12), pp.6999-7019, 2021.

[12] Guo, M.H., Xu, T.X., Liu, J.J., Liu, Z.N., Jiang, P.T., Mu, T.J., Zhang, S.H., Martin, R.R., Cheng, M.M. and Hu, S.M.,, "Attention mechanisms in computer vision: A survey," Computational visual media, 8(3), pp.331-368, 2022.

[13] Lin, T., Wang, Y., Liu, X. and Qiu, X.,, "A survey of transformers," AI open, 3, pp.111-132, 2022.

[14] Han, Z., Zhao, J., Leung, H., Ma, K.F. and Wang, W.,, "A review of deep learning models for time series prediction," IEEE Sensors Journal, 21(6), pp.7833-7848, 2019.

[15] Soydaner, D.,, "Attention mechanism in neural networks: where it comes and where it goes,"

Neural Computing and Applications, 34(16), pp.13371-13385, 2022.

[16] Ganesh, P., Chen, Y., Lou, X., Khan, M.A., Yang, Y., Sajjad, H., Nakov, P., Chen, D. and Winslett, M. , "Compressing large-scale transformer-based models: A case study on bert," Transactions of the Association for Computational Linguistics, 9, pp.1061-1080, 2021.

[17] Ahmed, S., Nielsen, I.E., Tripathi, A., Siddiqui, S., Ramachandran, R.P. and Rasool, G., "Transformers in time-series analysis: A tutorial," Circuits, Systems, and Signal Processing, 42(12), pp.7433-7466, 2023.

[18] Coon, W.G. and Ogg, M. , "Laying the foundation: Modern transformers for gold-standard sleep analysis," bioRxiv, pp.2024-01, 2024.

[19] Salur, M.U. and Aydin, I.,, "A novel hybrid deep learning model for sentiment classification," IEEE Access, 8, pp.58080-58093, 2020.

[20] Jena, B., Saxena, S., Nayak, G.K., Saba, L., Sharma, N. and Suri, J.S., "Artificial intelligence-based hybrid deep learning models for image classification: The first narrative review," Computers in Biology and Medicine, 137, p.104803, 2021.

[21] Rasool, M., Ismail, N.A., Boulila, W., Ammar, A., Samma, H., Yafooz, W.M. and Emara, A.H.M., , "A hybrid deep learning model for brain tumour classification," Entropy, 24(6), p.799, 2022.

[22] Pei, W.; Li, Y.; Siuly, S.; Wen, P., "A hybrid deep learning scheme for multi-channel sleep stage classification," Computers, Materials and Continua, 71(1), 2022, pp.889-905.

[23] Liu, Z.; Luo, S.; Lu, Y.; Zhang, Y.; Jiang, L.; Xiao, H., "Extracting multi-scale and salient features by MSE based U-structure and CBAM for sleep staging," IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31, 2022, pp.31-38.

[24] Zhao, C.; Li, J.; Guo, Y., "SleepContextNet: A temporal context network for automatic sleep staging based single-channel EEG," Computer Methods and Programs in Biomedicine, 220, 2022, p.106806.

[25] Zan, H. ; Yildiz, A, "Multi-task learning for arousal and sleep stage detection using fully convolutional networks," Journal of Neural Engineering 20, no. 5 (2023): 056034.

[26] Kong, G.; Li, C.; Peng, H.; Han, Z.; Qiao, H., "EEG-Based Sleep Stage Classification via Neural Architecture Search," IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31, 2023, pp.1075-1085.

[27] Li, C.; Qi, Y.; Ding, X.; Zhao, J.; Sang, T. ; Lee, M.,, "A deep learning method approach for sleep stage classification with eeg spectrogram," International Journal of Environmental Research and Public Health 19, no. 10 (2022): 6322.

[28] Hilal, A.M., Al-Rasheed, A., Alzahrani, J.S., Eltahir, M.M., Al Duhayyim, M., Salem, N.M., Yaseen, I. and Motwakel, A., "Competitive Multi-Verse Optimization with Deep Learning Based Sleep Stage Classification," Comput. Syst. Sci. Eng., 45(2), pp.1249, 2022.

[29] Quan, S.F.; Howard, B.V.; Iber, C.; Kiley, J.P.; Nieto, F.J.; O'Connor, G.T.; Rapoport, D.M.; Redline, S.; Robbins, J.; Samet, J.M. ; Wahl, P.W.,, "The sleep heart health study: design, rationale, and methods," Sleep 20, no. 12, 1997, 1077-1085.

[30] Yu, Y., Si, X., Hu, C. and Zhang, J., "A review of recurrent neural networks: LSTM cells and network architectures," Neural computation, 31(7), pp.1235-1270, 2019.

[31] Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J. and Sun, L., , "Transformers in time series: A survey," arXiv preprint arXiv:2202.07125, 2022.

[32] Banerjee, C., Mukherjee, T. and Pasiliao Jr, E.,, "An empirical study on generalizations of the ReLU activation function," Proceedings of the 2019 ACM Southeast Conference (pp. 164-167), 2019.

[33] Boursalie, O., Samavi, R. and Doyle, T.E., "Evaluation metrics for deep learning imputation models," In International Workshop on Health Intelligence (pp. 309-

322). Cham: Springer International Publishing, 2021.

[34] Basha, S.M. and Rajput, D.S., "Survey on evaluating the performance of machine learning algorithms: Past contributions and future roadmap," In Deep Learning and Parallel Computing Environment for Bioengineering Systems (pp. 153-164). Academic Press, 2019.

[35] Shrestha, A. and Mahmood, A., "Review of deep learning algorithms and architectures," IEEE access, 7, pp.53040-53065, 2019.

[36] Moradi, R., Berangi, R. and Minaei, B.,, "A survey of regularization strategies for deep models," Artificial Intelligence Review, 53(6), pp.3947-3986, 2020.

[37] Fernandez-Blanco, E.; Rivero, D. ; Pazos, A.,, "EEG signal processing with separable convolutional neural network for automatic scoring of sleeping stage," Neurocomputing, 410, 2020, pp.220-228.

[38] Pei, W.; Li, Y.; Wen, P.; Yang, F.; Ji, X., "An automatic method using MFCC features for sleep stage classification," Brain Informatics, 11(1), p.6, 2024.

[39] Zhang, D.; Sun, J.; She, Y.; Cui, Y.; Zeng, X.; Lu, L.; Tang, C.; Xu, N.; Chen, B. ; Qin, W., , "A two-branch trade-off neural network for balanced scoring sleep stages on multiple cohorts," Frontiers in Neuroscience, 17, 2023.

[40] Zhang, W.; Li, C.; Peng, H.; Qiao, H.; Chen, X, "CTCNet: A CNN Transformer capsule network for sleep stage classification," Measurement, 114157, 2024.