Mathematical Biosciences and Engineering

MBE, 20(9): 17018–17036.
DOI: 10.3934/mbe.2023759
Received: 20 April 2023
Revised: 06 July 2023
Accepted: 11 August 2023
Published: 29 August 2023

http://www.aimspress.com/journal/MBE

Research article

A convolutional neural network-based decision support system for neonatal quiet sleep detection

Saadullah Farooq Abbasi[1], Qammer Hussain Abbasi[2], Faisal Saeed[3] and Norah Saleh Alghamdi[4,*]

[1] Department of Biomedical Engineering, Riphah International University, Islamabad 44000, Pakistan
[2] James Watt School of Engineering, University of Glasgow, Glasgow, G4 0PE, United Kingdom
[3] DAAI Research Group, Department of Computing and Data Science, School of Computing and Digital Technology, Birmingham City University, Birmingham B4 7XG, UK
[4] Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

* **Correspondence:** Email: nosalghamdi@pnu.edu.sa.

Abstract: Sleep plays an important role in neonatal brain and physical development, making its detection and characterization important for assessing early-stage development. In this study, we propose an automatic and computationally efficient algorithm to detect neonatal quiet sleep (QS) using a convolutional neural network (CNN). Our study used 38-hours of electroencephalography (EEG) recordings, collected from 19 neonates at Fudan Children's Hospital in Shanghai, China (Approval No. (2020) 22). To train and test the CNN, we extracted 12 prominent time and frequency domain features from 9 bipolar EEG channels. The CNN architecture comprised two convolutional layers with pooling and rectified linear unit (ReLU) activation. Additionally, a smoothing filter was applied to hold the sleep stage for 3 minutes. Through performance testing, our proposed method achieved impressive results, with 94.07% accuracy, 89.70% sensitivity, 94.40% specificity, 79.82% F1-score and a 0.74 kappa coefficient when compared to human expert annotations. A notable advantage of our approach is its computational efficiency, with the entire training and testing process requiring only 7.97 seconds. The proposed algorithm has been validated using leave one subject out (LOSO) validation, which demonstrates its consistent performance across a diverse range of neonates. Our findings highlight the potential of our algorithm for real-time neonatal sleep stage classification, offering a fast and cost-effective solution. This research opens avenues for further investigations in early-stage development monitoring and the assessment of neonatal health.

Keywords: neonatal sleep; convolutional neural network; electroencephalography; polysomnography; biomedical engineering

1. Introduction

Sleep is an important aspect of maintaining good health in humans, especially for neonates, as it plays a vital role in their brain and physical development. Neonates spend 80% of their time sleeping in a resting state. Therefore, monitoring neonatal sleep in a neonatal intensive care unit (NICU) is essential. Neonatal sleep is commonly classified into three stages: active sleep (AS), quiet sleep (QS) and intermediate sleep (IS). Currently, polysomnography (PSG) is the leading technique for classifying neonatal sleep stages. PSG involves continuous monitoring of overnight sleep using multiple bio-physiological signals, including electrocardiography (ECG), electromyography (EMG), electrooculography (EOG), electroencephalography (EEG) and respiration. Professional neurologists then analyze and classify these signals into different sleep stages.

PSG is effective, but it is costly and time-consuming, and it necessitates expert neurologists for manual analysis. In recent years, various research initiatives have explored the feasibility of sleep classification utilizing PSG signals, with EEG emerging as the most dependable signal for both adults [1,2] and infants [3–5]. The first human EEG was recorded by Hans Berger in 1924 [6]. The electrical activity of the brain takes place through electrical impulses and can be ascertained from the scalp. Neurologists have identified EEG patterns for sleep-wake cycles starting from 30 weeks postmenstrual age [7]. Loomis et al. published the first study on human sleep patterns using EEG in 1937 [8]. Since then, numerous algorithms have been suggested for adult sleep staging employing machine learning [9–12] and deep learning [13–17].

Neonates and adults have different sleep states. The neonatal sleep stages are mainly divided into two main types: active sleep (AS) and quiet sleep (QS). Frequently, sleep cannot be assessed as being either AS or QS and is termed intermediate sleep, which is, in fact, a "transitional sleep" [18]. Transitional sleep does not meet the requirements for AS and QS, but it happens for a short time during the transitional periods between these clearly defined states. Meanwhile, adult sleep stages are two-pronged: rapid eye movement (REM) and non-rapid eye movement (NREM). NREM is then further classified into NREM1, NREM2, NREM3 and NREM4. Recently, NREM3 and NREM4 have been combined to form the NREM3 stage. The primary differentiation is the sleep time: Neonates spend 17–18 hours (70%) sleeping in a resting state, whereas adults spend 7–9 hours (16–29%) in a sleeping state. Neonates do not exhibit circadian rhythms, whereas an adult's EEG contains circadian rhythms which help them sleep during the night. A secondary contrast is sleep spindles, which are pertinent for memory consolidation, yet they are not apparent in neonates. Sleep spindles develop after 42 weeks of gestational age (GA) [19,20]. Another difference is the sleep cycle time: The neonate sleep cycle completes in 50–60 minutes, whereas the adult sleep cycle averages 80–110 minutes. Another difference is the percentage of sleep in REM and non-REM sleep. Figure 1 shows the distribution of AS and QS with respect to gestational age (GA). At 28 weeks of GA, 80% of the total sleep time (TST) constitutes AS, whereas 18% is in QS. The remaining 2% is transitional sleep or IS, which is the transitional sleep from AS to QS and vice versa. The percentage of AS decreases with an increase in GA, whereas QS increases. Finally, at 40 weeks of GA, the percentage of AS is 57%, whereas for QS, it is 32%, and the remaining 11% is transitional sleep [21].

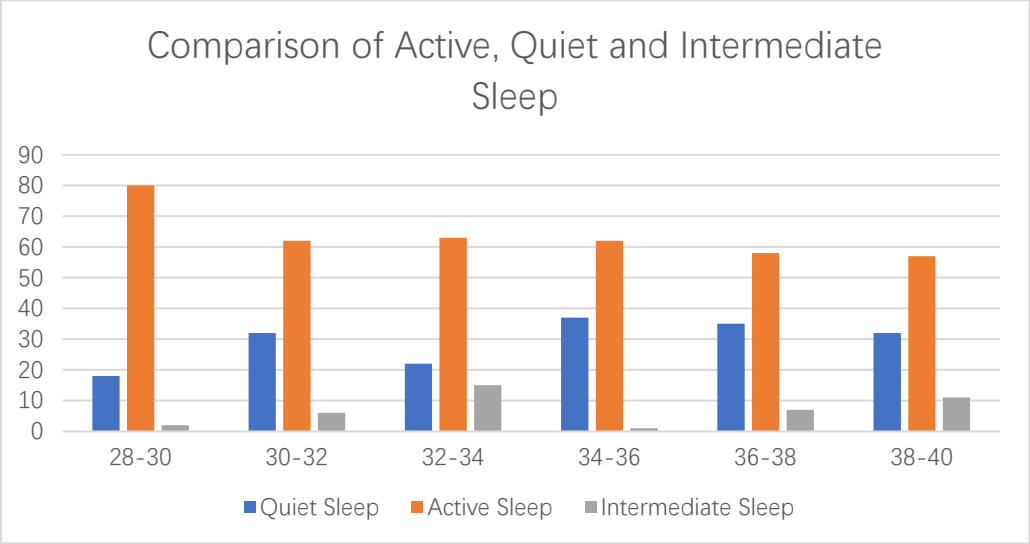

Figure 1. Percentages of AS and QS with respect to gestational age.

This study proposes a novel algorithm based on convolutional neural networks (CNNs) for detecting neonatal quiet sleep (QS). By leveraging the power of deep learning and EEG signal analysis, the proposed model offers an automated and computationally efficient solution for accurately classifying neonatal sleep stages. The motivation behind our research lies in the potential impact it can have on neonatal care. An automated, fast, inexpensive and reliable sleep stage classification system can assist healthcare professionals in monitoring neonatal sleep patterns, evaluating brain maturation and identifying any deviations that may indicate underlying health issues. Moreover, by reducing the reliance on manual scoring methods, our proposed algorithm can improve efficiency and consistency in neonatal sleep stage assessment, ultimately enhancing the quality of care provided to newborns. The study holds promise for the development of real-time clinical applications that can benefit healthcare professionals, neonatal care units and families by facilitating accurate sleep stage monitoring and early-stage development assessment. The main contributions of this paper are given below:

- To the best of our knowledge, to date, there is no public dataset available online for neonatal sleep classification. Therefore, we used neonatal EEG data from Fudan children's hospital Shanghai, China, involving multiple experienced neurologists for the annotation.
- There is pre-processing of neonatal EEG using the latest signal processing tools and comparison with the already established schemes.
- We design a novel dedicated CNN model to detect neonatal QS.

The rest of the paper is arranged in the following manner: Section 2 presents the related work in the field of neonatal sleep classification, Section 3 proposes the methodology, Section 4 presents the results, and Section 5 discusses the overall scheme. Finally, Section 6 concludes the paper.

2. Related work

In 1937, Loomis et al. [8] put forth the preliminary EEG-based study of human sleep patterns. After the introduction of high-performance supercomputers, machine/deep learning algorithms grow exponentially in the biomedical field, classification, prediction, emotion recognition [22] and signal

processing. In 2001, Turnbull et al. [23] developed an automatic sleep state detection algorithm using discrete wavelet transform (DWT). Tracé alternant (TA) detection using DWT was carried out with six healthy neonates, within 3 days after birth. All 20 segments of TA were successfully detected using the proposed scheme. The DWT-based algorithm detected TA effectively; however, QS also includes trace discontinu (TD) and high voltage slow (HVS) signals. The proposed algorithm performs well for TA detection but cannot be used for entire QS detection over a wider range of neonates.

Certain developmental changes can only be observed during QS. For this reason, Piryatinska et al. [24] proposed a supervised classical feature-based QS detection; 9 time and frequency domain features were extracted from 30-second neonatal EEG. Finally, support vector machine (SVM) with radial basis function (RBF) was applied for QS detection. The proposed algorithm achieved a mean kappa of 0.7. To further enhance the performance, multiple algorithms have been proposed for machine/deep learning-based QS detection. Koolen et al. [4] proposed an algorithm based on a greedy algorithm; 57 EEG features were extracted from 8-channel EEG using 10-minute epochs. However, an increased set of features can cause over-fitting, so a greedy algorithm was used to reduce the set of features. Finally, SVM was used for classification. The results demonstrated that the proposed network worked well with the greedy algorithm, reaching an accuracy of 85%. In 2017 [3], Dereymaeker et al. proposed an algorithm for automatic sleep stage classification using cluster-based adaptive sleep staging (CLASS); 9 EEG features were extracted from 89 preterm neonates. The CLASS-based algorithm can reach a ROC curve of 97% for preterm neonates. It outperformed existing algorithms for QS detection in preterm neonates but cannot be used as an overarching configuration for every neonate.

Extracting complex features from multichannel EEG is a time-consuming process and may not be considered for real-time applications; therefore, a deep learning method using deep CNN was proposed by Ansari et al. in 2018 [25]. A total of 97 EEG recordings were collected, among which 54 were used for training, and 43 were used for testing. EEG data were downsampled to 30 samples/sec to decrease the complexity of the neural network. Seven convolution layers were used to train the neural network. The network can reach a mean area under the curve of 92%. The proposed network outperformed existing networks for QS detection, yet there is still some need for improvement. An unsupervised QS-detection algorithm was proposed by [26] in 2019. In this method, the multichannel EEG data is processed using a comprehensive procedure. Initially, the EEG data undergoes a tensorization process utilizing multiscale entropy, followed by factorization through canonical polyadic decomposition (CPD) to yield a collection of rank-1 tensors. Through autocorrelation analysis, the most appropriate CPD component associated with sleep stages is automatically identified. Subsequently, the detection of QS intervals is accomplished by applying k-means clustering to the smoothed temporal pattern extracted from the selected component. The algorithm achieved an accuracy of 79%. However, it used core tensors; therefore, tensor reduction was not possible. To classify all three stages, Fraiwan et al. proposed an algorithm using deep autoencoders and a long short-term memory learning system in 2017 [27] and 2020 [28], respectively. The algorithms worked well for three stages but less for QS detection.

Table 1 shows the literature survey of the already proposed neonatal QS detection algorithms, containing accuracy and their limitations.

Table 1. Literature review for automatic quiet sleep detection.

Reference	Year	Algorithm	Stages	Evaluation	Limitation
[23]	2001	Discrete wavelet transforms	Tracé alternant	Acc = 100%	Tracé discontinu / High voltage slow signals cannot be classified
[24]	2009	FBD-SVM	QS-detection	Acc = 89%	Less accuracy
[4]	2017	Greedy algorithm	QS-detection	Acc = 85%	Increased number of features increases the overfitting
[3]	2017	CLASS	QS-detection	AUC > 90%	Only classification for preterm infants
[25]	2018	CNN	QS-detection	AUC > 92%	Complex network
[26]	2019	Multiscale entropy tensor decomposition	QS-detection	Acc = 79%	Utilization of core tensor for classification
[27]	2017	Autoencoders	QS-detection	Acc = 77.6%	Less accuracy
[28]	2020	LSTM	QS-detection	Acc = 94.6%	Complex network

3. Materials and methods

This section will explain the dataset, feature extraction and classification algorithm in detail. Figure 2 presents the block diagram of the proposed study.

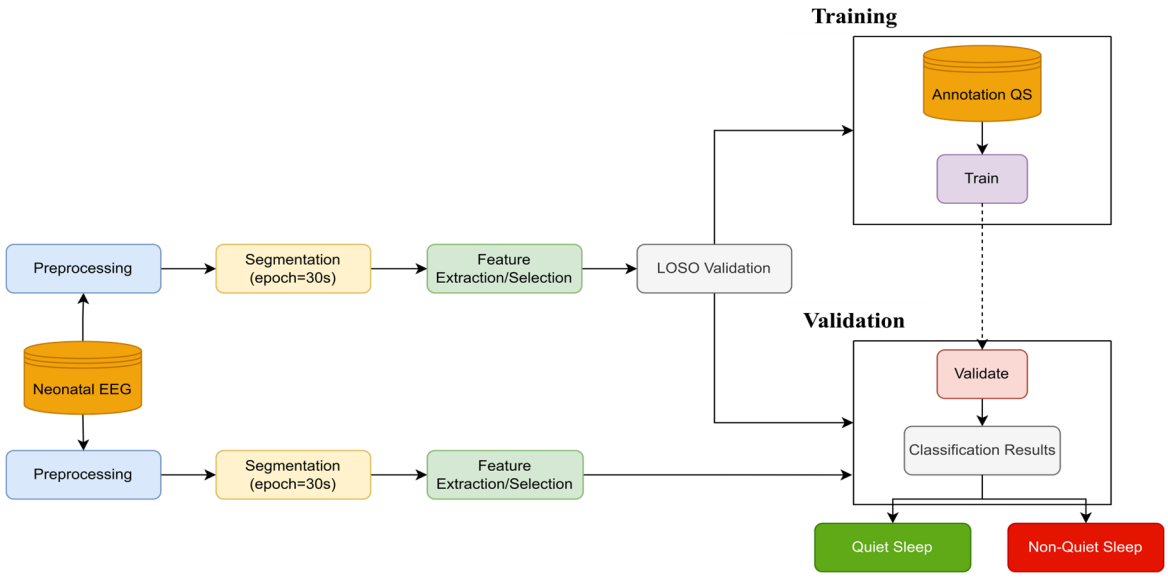

Figure 2. Overview of the proposed study.

3.1. Dataset and feature extraction

Neonatal data from 19 preterm and term neonates were used for this experiment. The neonatal VEEG data was collected from a neonatal intensive care unit (NICU) at Fudan Children's Hospital Shanghai, China [29–34]. Approval was granted by both the Internal Ethics Committee for Neonatal Experiments of the Children's Hospital Psychological Ethical and the Test Committee of Fudan University, Shanghai, China (Approval No. (2020) 22). For data extraction and annotation, a

NicoletOne EEG IoT-based system was used. The video-EEG (VEEG) data was extracted from 9 bipolar EEG channels (Fp1, Fp2, T3, C3, Cz, C4, T4, O1 and O2) with a sampling frequency of 500 Hz. A concise summary of the dataset is provided in Table 2. From the table, it is apparent that the neonatal population under investigation does not exhibit any significant pathology or medical conditions that may impact their brain activity or sleep patterns.

Table 2. Details of the subjects used for neonatal sleep stage classification.

#	Gender	HD	CC	W	H	BP	P	R	GA (wk^{+d})	Reason for admission
1	Girl	32.8	30.6	2.6	48	68/38	130	35	34^{+0}	Preterm
2	Boy	36	40	4.5	56	70/40	150	45	39^{+0}	Fever
3	Girl	33	30	2.6	49	64/45	122	70	36^{+4}	Premature/Anhelation
4	Boy	35	31.5	3.3	50	74/39	135	50	39^{+2}	Cyanosis
5	Boy	34	33	2.9	50	71/40	118	50	37^{+1}	Jaundice
6	Girl	29	28	2.2	45	65/40	150	45	35^{+3}	Premature
7	Girl	33	32.5	2.3	48	77/52	146	44	34^{+6}	Premature & emesis
8	Boy	33	29	2.8	46	66/36	140	47	38^{+5}	Hypoglycemia
9	Boy	35	34	3.7	50	70/40	150	45	37^{+5}	Fever
10	Boy	36	33	3.0	47	66/39	146	44	38^{+6}	Fever
11	Girl	32	31	2.7	49	70/40	140	40	40^{+4}	Jaundice
12	Girl	32	33	2.0	48	70/40	146	42	34^{+1}	Jaundice
13	Boy	30.5	28	2.2	46	70/40	140	52	35^{+3}	Anhelation
14	Girl	33	32	3.2	47	70/42	146	42	35^{+0}	Preterm
15	Boy	28.5	26	1.6	41	68/33	160	55	31^{+3}	Preterm/Anhelation
16	Girl	34	32	3.2	50	70/40	142	52	40^{+5}	Jaundice
17	Boy	33	33	3.1	50	70/40	140	40	36^{+1}	Per-term/Jaundice
18	Boy	34	33	3.5	50	70/40	140	40	39^{+2}	Jaundice
19	Girl	34	33	2.9	50	70/40	135	40	38^{+0}	Jaundice

Note: HD: Head circumference, CC: Chest circumference, W: Weight (kg), H: Height (cm), BP: Blood pressure, P: Pulse per minute, R: Respiratory per minute, GA: Gestational age (weeks).

During recording and transmission, it is common for EEG recordings to become contaminated with various types of noise and artifacts. There are typically two main types of noise in EEG recordings: powerline noise and baseline noise. Powerline noise is caused by 50 or 60 Hz interference from electrical wiring, while baseline noise arises from poor contact between the electrodes and the scalp. In addition to noise, artifacts in EEG data can also be present, which refer to signals that are not of cerebral origin and can arise from various physiological sources, such as muscle activity, eye movements and ECG signals. These unwanted signals must be removed before performing feature extraction and classification. Figure 3 shows the 30-second EEG recordings of QS and non-QS stages.

To address this issue, a finite impulse response (FIR) filter was utilized with cut-off frequencies of 0.3–35 Hz to remove the noise and artifacts. After applying the filter, 30-second epochs were separated, and each epoch was labeled with its respective state by professional neurologists using visual labeling. The epochs labeled as "AS" (active sleep) and "awake" were combined to form non-QS (quiet sleep) epochs. Epochs labeled as "artifacts" were manually removed from a total of 4560 segments.

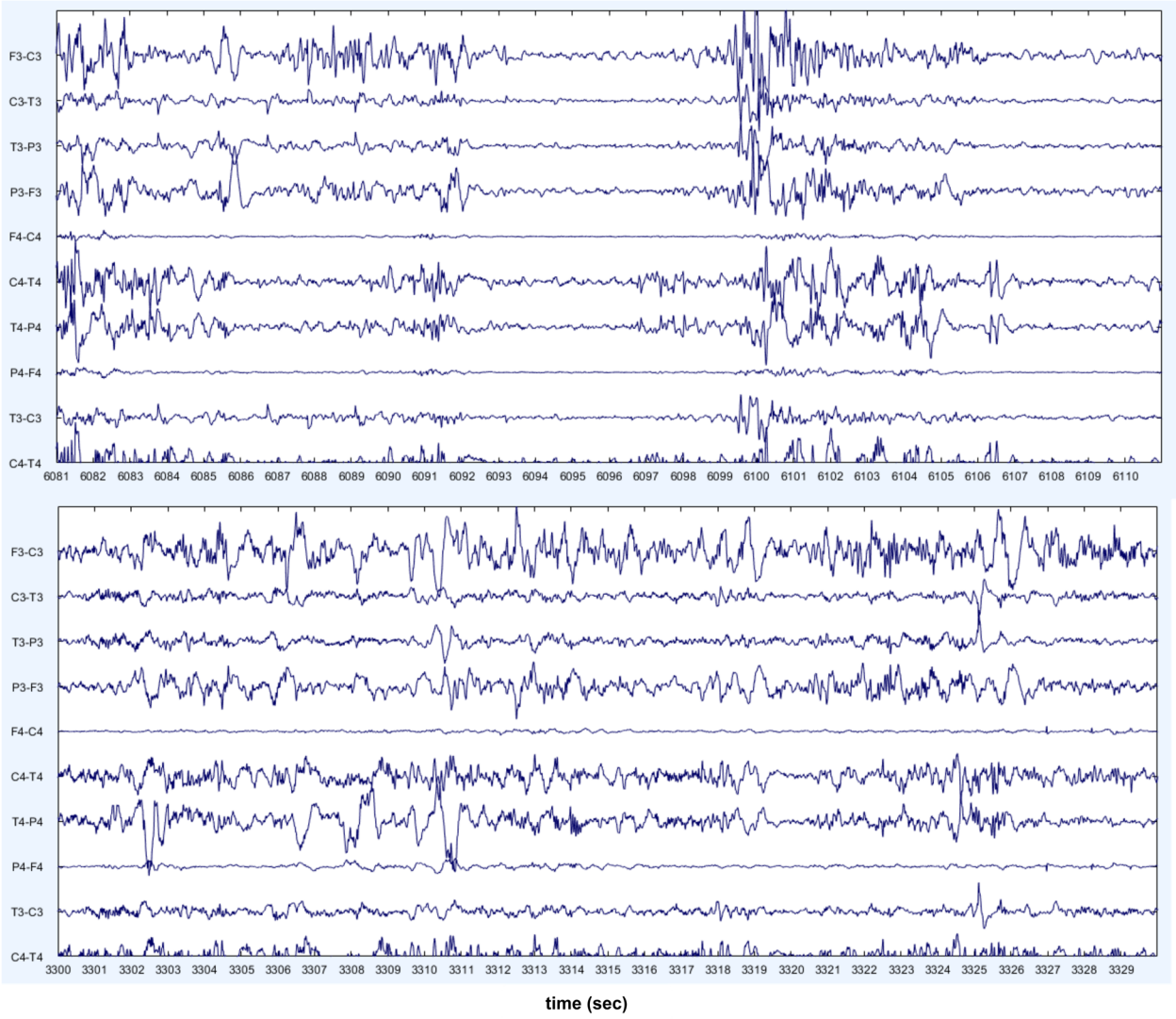

Figure 3. Sample EEG recordings for neonatal QS (upper) and non-QS (lower).

Table 3. Features extracted from neonatal sleep EEG epochs.

Time domain features		Frequency domain features
Mean	Skewness	Spectral power
Median	Kurtosis	Relative power
Peak-to-Peak distance	Standard deviation	Peak frequency
Number of peaks	Variance	Mean frequency of:
Maxima	Correlation	• Alpha band
Minima	Activity	• Theta band
Number of zero crossings	Mobility	• Delta band
Zero crossing rate	Complexity	• Beta band
Frequency domain	Entropy	

Feature extraction is about dimensionality reduction, which transforms large raw data variables into numerical features. The proposed study extracted 25 features from bipolar neonatal EEG recordings. These features are mainly categorized into two types: time domain and frequency domain.

Time domain features of EEG refer to measures of the signal that describe the characteristics of the electrical activity over time, whereas frequency domain features of EEG refer to the analysis of the EEG signal in the frequency domain. These frequency domain features can provide information about the power, frequency content and organization of the electrical activity in the brain. Further classification of these features is given in Table 3.

After extracting features from the EEG data, various feature selection techniques were utilized to reduce the complexity and dimensionality. Two methods, namely, the correlation-based feature selection technique and manual feature engineering, were employed to select the twelve most significant bipolar EEG features. These features were classified into two categories: eight time-domain features and four frequency-domain features. The time-domain features included mean, median, skewness, kurtosis, minimum, maximum, standard deviation and variance, while the frequency-domain features consisted of the mean frequencies of alpha, beta, theta and delta bands.

3.2. Methodology

In this study, a CNN has been designed for automatic quiet sleep detection. As an artificial neural network (ANN), it is composed of interleaved Convolutional (Conv), Linear and Pooling layers. The Conv layer functions as an expanded version of the FIR filter bank. Since the hidden layers are nonlinear, a nonlinear operator (ReLU) has been employed after each Conv layer. ReLU operates as a conventional half-wave rectifier that retains positive values while replacing negative values with zero.

The CNN proposed in this study involves three main operations: convolution, pooling and activation function. The convolution operation comprises sliding a filter over the input data to calculate the dot product between the filter and the input region. Mathematically, the convolution operation can be expressed as:

$$y(i,j) = f(x(i{:}\, i+k-1, j{:}\, j+k-1) * w + b) \tag{1}$$

Here, $y(i,j)$ is the output of the convolution operation, $x(i{:}\, i+k-1, j{:}\, j+k-1)$ is the input at the region (i,j), w is the weight, b is the bias, and f is the activation function. The output of the convolution operation is linear; therefore, a rectified linear unit (ReLU) is utilized to introduce nonlinearity. ReLU sets negative values to zero, thus eliminating noise and irrelevant information in the feature maps, making the features more distinct and easier to learn by the following layers. Mathematically, ReLU can be expressed as

$$f(x) = \max(0, x). \tag{2}$$

The pooling operation is employed to decrease the dimensionality of the feature maps by aggregating neighboring values. In this study, maximum pooling is used, which is expressed mathematically as

$$y(i,j) = \max\big(x(is{:}\, is+k-1, js{:}\, js+k-1)\big) \tag{3}$$

Here, $y(i,j)$ represents the output of the pooling operation at location (i,j), $x(is{:}\, is+k-1, js{:}\, js+k-1)$ is the local region of the input, k is the size of the pooling window, s is the stride, and max is the maximum function. It is noteworthy that these layers have been chosen by trial and error. The model that yields the highest mean accuracy and kappa has been reported.

The proposed CNN is composed of two Conv layers, two ReLU operators, 1 max pooling layer and 2 densely connected layers. Figure 4 shows the architecture of the proposed CNN network for QS classification. The input is a 9-channel EEG segment, each containing 12 features. The first Conv layer has 50 kernels of size 1 × 9, where the mode operates on channels. This process is followed by the ReLU operator for rectification. This proceeds to the second Conv layer with 30 kernels of size 1 × 4, followed by another ReLU and a max pooling layer of size 2 × 2. Finally, a flattening layer was used to arrange the feature map into a single vector. The features are then fed into a two-layer multilayer perceptron (MLP) for classification. The learning rate (LR) is set to 0.001 with 50 epochs. In addition to CNN, the proposed algorithm uses a smoothing filter as a post-processing step. From the literature, it is believed that a neonate spends at least 3 minutes in one sleep stage. For this reason, this filter halts the sleep state for 6 epochs, i.e., 6 × 30 seconds = 3 minutes.

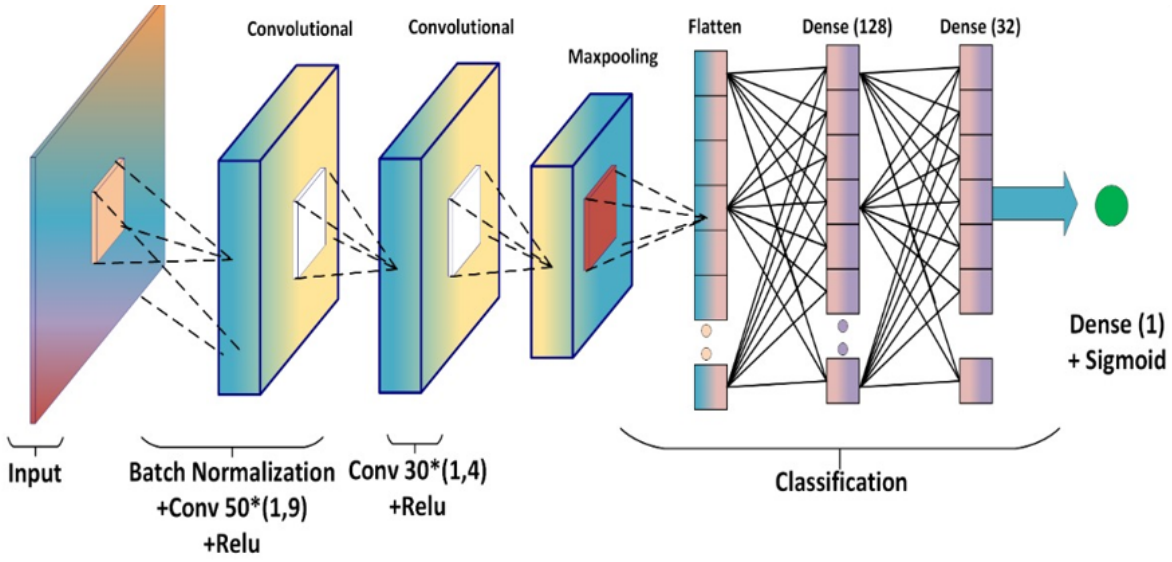

Figure 4. CNN architecture for QS detection.

3.2.1. Evaluation

Multiple evaluation parameters have been used to assess the performance of the proposed scheme. These metrics are mainly categorized into five types:

3.2.1.1. Accuracy

Accuracy is an important parameter to assess the performance of a classification model. It measures the percentage of correct predictions made by the classification model. Specifically, accuracy is defined as follows:

$$Accuracy = \frac{T_P+T_N}{T_P+T_N+F_P+F_N} \tag{4}$$

where T_P is true positives, T_N is true negatives, F_P is false positives, and F_N is false negatives.

3.2.1.2. Kappa

The kappa statistic is a measure of agreement between the classifications made by a classifier and the ground truth labels. It takes into account the expected accuracy of a random classifier and allows for control of chance agreement between the two sets of data. The classes of this study are unbalanced; therefore, kappa is an important parameter to validate the efficacy of the proposed classifier. Mathematically, Kappa is given as

$$Kappa = \frac{P_{obs} - P_{ch}}{1 - P_{ch}} \tag{5}$$

The value of P_{obs} in Cohen's kappa equation represents the proportion of observed agreement between two raters or classifiers. On the other hand, P_{ch} is the proportion of agreement that would be expected by chance alone. P_{ch} is mathematically given as

$$P_{ch} = (P_a * P_b) + ((1 - P_a) * (1 - P_b)) \tag{6}$$

where P_a and P_b are the observations that were classified by each rater or classifier.

3.2.1.3. Sensitivity

Sensitivity is a measure of the ability of the classification model to predict the QS class. Mathematically,

$$Sensitivity = \frac{T_P}{T_P + F_N}. \tag{7}$$

3.2.1.4. Specificity

Specificity is a measure of the ability of the classification model to predict the non-QS class. Mathematically,

$$Specificity = \frac{T_N}{T_N + F_P}. \tag{8}$$

3.2.1.5. F1-score

Another important parameter to assess the classification performance of unbalanced data is F1-score. It is the measure of the harmonic mean of precision (P) and recall (R). Mathematically,

$$F1 - score = 2 * \frac{P * R}{P + R}. \tag{9}$$

As the data is imbalanced, the key parameter to evaluate the proposed classifier is mean kappa. As this study contains an imbalanced dataset, the minority class may have a disproportionately smaller number of instances compared to the majority class, leading to problems in evaluating the classifier's performance. For instance, a model that predicts the majority class for every sample may yield high accuracy but is not practical. Cohen's Kappa considers the probability of chance agreement between two evaluators and produces a score between -1 and 1. A score of 0 indicates that the classifier's

performance is no better than random chance, whereas a score of 1 represents a perfect match between the classifier's predictions and the true labels.

The performance of the classifier was first evaluated using a leave-one-subject-out (LOSO) cross-validation approach, where the model was trained using data from $n-1$ neonates and evaluated on the data of the one neonate that was held out. This process was repeated for all n folds. During training, a portion of the data from each training fold in the LOSO cross-validation was reserved for inner validation to avoid overfitting. Finally, stratified mean has been reported to summarize the results of the classifier's performance evaluation.

4. Results

An Intel Core i5-8400 with 16 GB RAM and GTX 1050ti was used to train and test all networks. Keras and TensorFlow were used to implement the proposed neural network, and MATLAB 2019b was used to obtain the features. The proposed CNN can accurately detect neonatal QS with 94 ± 1.2% accuracy, as shown in Table 4, which displays the performance parameters, including kappa, sensitivity, specificity and accuracy.

Table 4. Performance metrics of the proposed CNN.

	Accuracy	Kappa	Sensitivity	Specificity	F1-score	Positive predictive value	Negative predictive Value
Proposed CNN	94.07%	0.74	89.70%	94.40%	79.82%	82.90%	95.98%

In Figure 5, the test performance of the proposed network and the current state-of-the-art algorithm are compared. Kappa and other parameters were computed from test recordings to determine the effectiveness of the proposed CNN. The results of the proposed CNN were compared with other advanced methods to verify its efficacy. Figure 5 shows that the proposed approach outperformed the existing state-of-the-art networks for detecting QS.

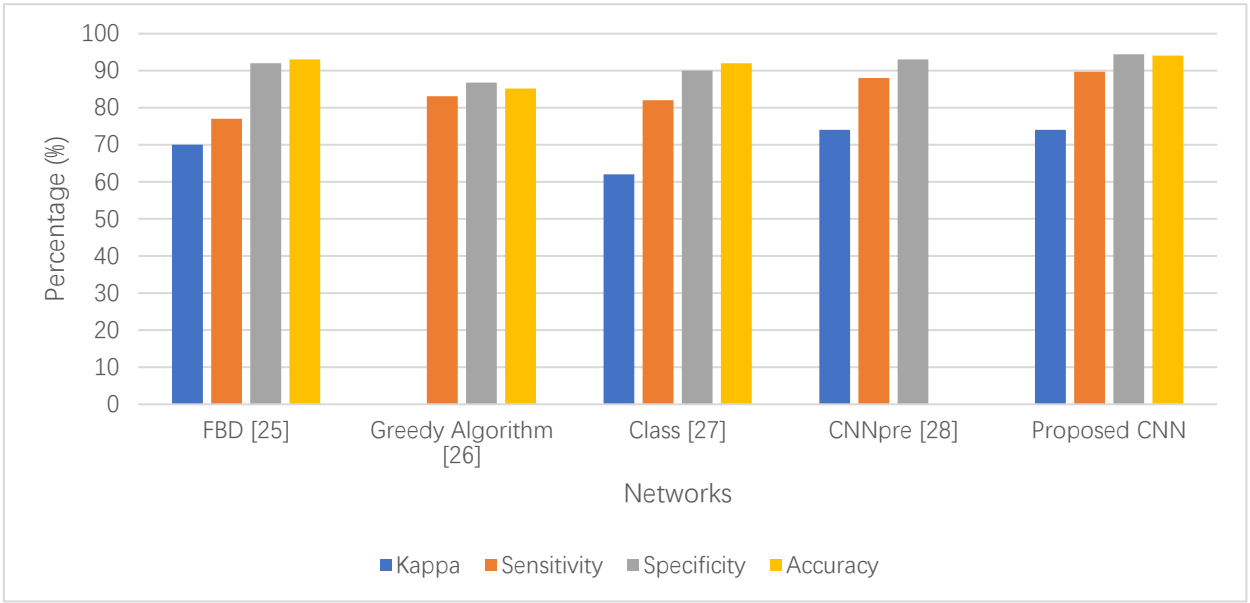

Figure 5. Performance comparison of the proposed algorithm with existing networks.

This study developed a classification model that can accurately identify neonatal sleep stages. It also provides additional evidence that brain activity varies during quiet and non-quiet sleep. It has been found that the accuracy of our sleep stage classifier was affected by postmenstrual age (PMA), which could be due to the maturation of the brain. Although sleep patterns begin to develop at 27 weeks, it is not until around 31 weeks of PMA that sleep-wake cycles can be distinguished [35]. Since this study did not include data from the extreme preterm group, our classifier may be biased toward other age groups that are more represented. Figure 6 shows the receiver operating characteristics curve of the proposed scheme. In this study, the distribution of subjects across the groups is as follows: For gestational age (GA) greater than 37 weeks, we have a total of 10 subjects. Conversely, for subjects with a gestational age below 37 weeks, there are 9 individuals. From the figure, it can be seen that the proposed scheme performs better for neonates with GA > 37 weeks.

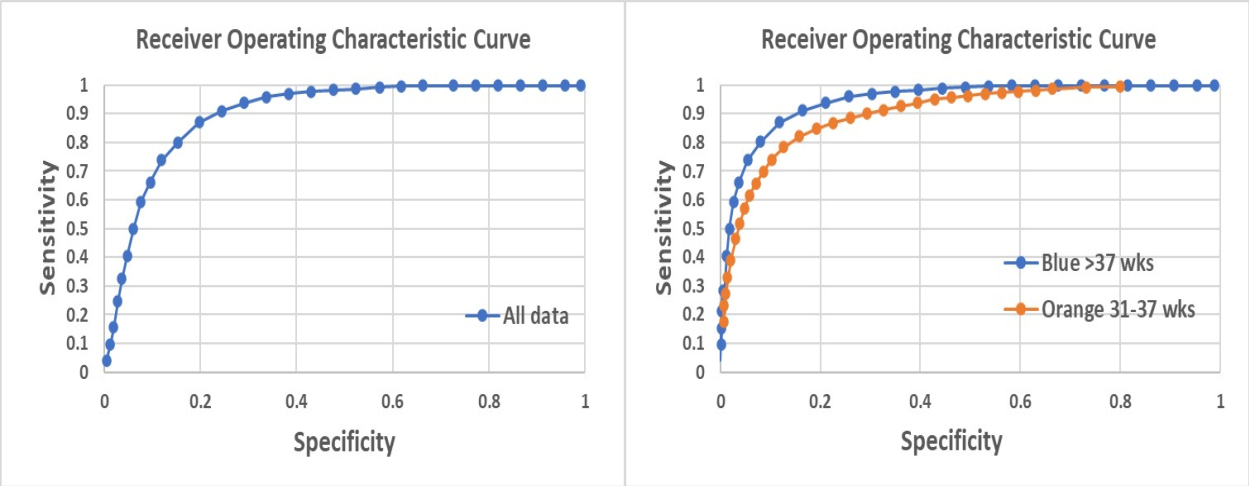

Figure 6. Receiver Operating Characteristic curves of the proposed scheme.

A hypnogram shows the best results for classification. For this reason, the hypnogram of the proposed algorithm is presented and compared with the neurologist's annotation. Figure 7 shows the histogram of the professional neurologist, whereas Figure 8 shows the result of the proposed CNN.

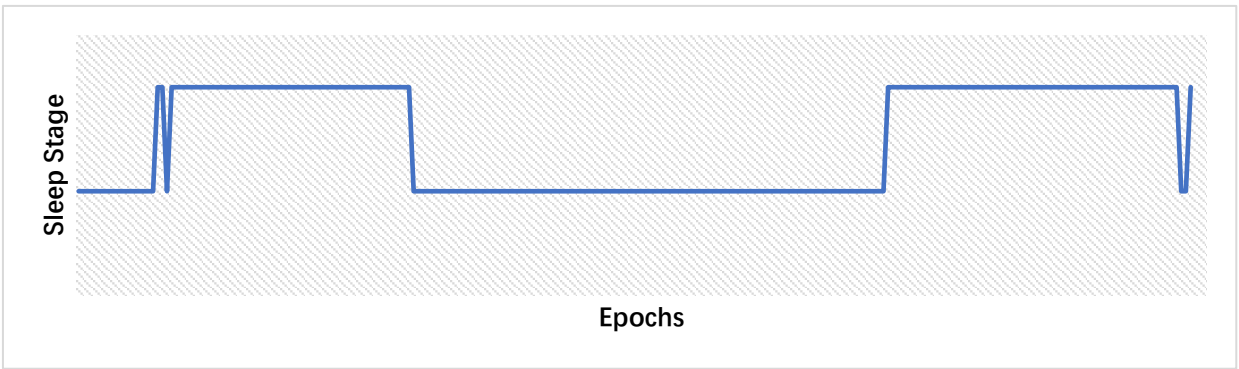

Figure 7. Professional neurologist hypnogram.

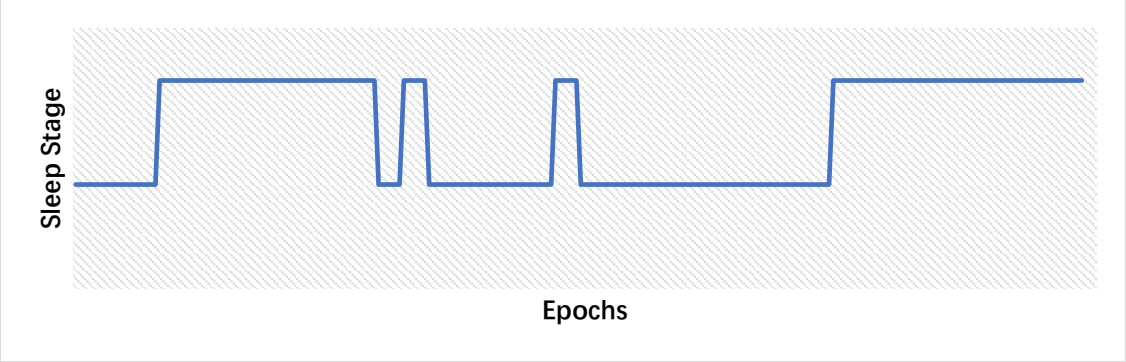

Figure 8. The proposed CNN hypnogram.

Multichannel EEG extraction devices may not be available in every NICU. As a result, Figure 9 displays the outcomes of the proposed network when varying numbers of EEG channels are used. In each instance, the CNN was trained using distinct random initializations, and the most effective network is displayed in Figure 9. The proposed CNN network outperforms existing methods in each scenario. However, as the number of channels is reduced, accuracy declines.

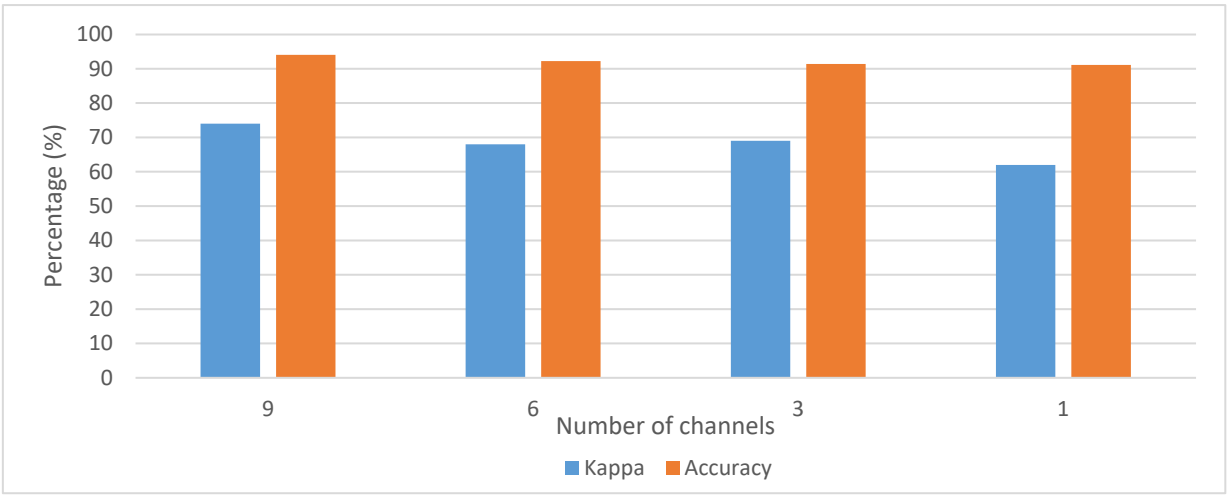

Figure 9. Performance of CNN with different numbers of EEG channels.

There is a belief that machine learning algorithms work better with hand-crafted features. To demonstrate the efficiency of the suggested approach versus machine learning models, various algorithms were utilized and compared. Table 5 displays the comparison between the suggested approach and machine learning techniques.

Table 5. Performance comparison of CNN vs machine learning algorithms.

Models	Kappa	Accuracy
KNN (K = 10)	0.61	89.77
SVM	0.65	91.89
Decision tree	0.55	85.10
Proposed CNN	0.74	94.07

According to the literature [31,36], a subject spends at least three minutes in a specific sleep stage. Therefore, a smoothing filter is used to halt a specific sleep stage for 3 minutes, i.e., 6 epochs. This increases the accuracy of the proposed scheme by 2%. Table 6 shows the results of the proposed CNN model with and without a smoothing filter. However, this may be considered a drawback because this is a post-processing technique that limits an application that is to work in a real-time environment.

Table 6. Performance of CNN with smoothing filter.

	Kappa	Accuracy
With smoothing filter	0.74	94.07
Without smoothing filter	0.71	91.79

In this fast and emerging world, neural networks are used for easy and fast implementation. Computational time is one of the main aspects in determining the quality of the proposed model. To validate the minimum computational complexity of the proposed study, Figure 10 shows a bar graph illustrating the mean evaluation times for 2-hour EEG segments. The graph clearly demonstrates that our study surpasses existing schemes for neonatal QS detection in terms of computational efficiency.

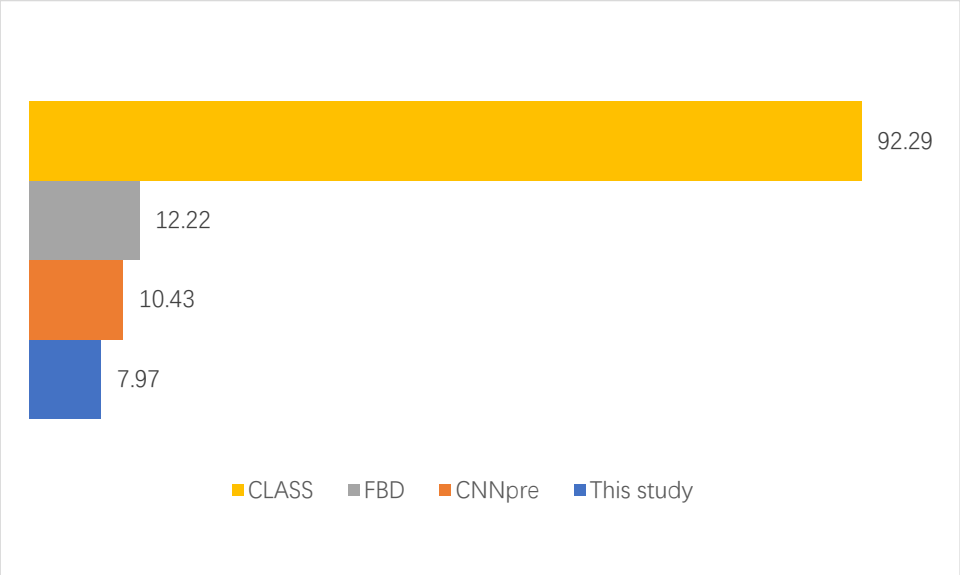

Figure 10. Mean computational time (seconds) for 2-hour EEG data classification.

5. Discussion

This study established a model for detecting neonatal QS. As stated in the introduction, certain maturational changes can only be observed during QS, making it critical to classify this sleep stage accurately. The proposed model employed a CNN model to classify QS. Indigenously developed features from 9 bipolar EEG channels were used to train the classification model. The model achieved an accuracy of 94.07% in identifying neonatal QS.

While the performance is better as compared to the existing algorithms, this model possesses some important benefits. It is computationally inexpensive and uses only two convolutional layers for classification. For 2-hour data classification, the proposed scheme only requires 7.97 secs. This makes

it an efficient algorithm for real-time clinical applications. Also, the key challenge to classify the neonatal sleep stage is the fast alterations in the EEG patterns during sleep. The proposed network classified 30-second epochs with high accuracy.

Our findings, which are consistent with previous research, demonstrate that the complexity of neural activity is contingent upon the sleep stage [36,37]. By creating a classification model that accurately detects neonatal sleep stages, our study provides further support for the notion that brain dynamics differ between quiet and nonquiet sleep patterns. Notably, this study observed that the performance of our sleep stage classifier was impacted by postmenstrual age (PMA), which may be attributed to maturation effects. Although sleep state organization commences at 27 weeks, it is not until approximately 31 weeks of PMA that sleep-wake cycles become distinguishable. Our study did not include data from the youngest age group, resulting in an underrepresentation of this population in the study. Therefore, the classifier may be biased toward other age groups that are more represented. Alternatively, it is possible to develop separate classifiers for different age groups, which may outperform the presented classifier.

Typically, CNN necessitates a larger dataset for classification. However, effective feature extraction, along with the best model, results in successful neonatal sleep classification. Twelve EEG features from neonatal EEG were employed for both training and testing. Additionally, a post-processing smoothing filter was employed to pause the sleep stage for 3 minutes following classification. This resulted in a 2% improvement in accuracy. While this smoothing filter can improve the proposed method, there may be some drawbacks associated with the post-processing step.

The main limitation lies in multiple aspects. First, CNN performance can be increased by a larger dataset. In the proposed model, EEG data was extracted from 19 neonates. In the literature, it has been proved that by increasing the dataset, the performance of a CNN-based classifier can be improved. By having a larger dataset, the CNN may classify sleep stages from raw EEG. Another drawback of the proposed scheme is the use of wearable electrodes. These wearable electrodes can affect the sleep quality of the fragile neonate.

6. Conclusions

This study presented a CNN model for neonatal quiet sleep (QS) detection, achieving an impressive accuracy of 94% with a mean kappa coefficient of 0.74. Our model leveraged twelve prominent features extracted from nine bipolar EEG channels and employed a CNN architecture comprising two convolutional layers, pooling layers and ReLU activation. Additionally, a smoothing filter was applied to enhance the stability of sleep stage classification over six epochs. The results of our study demonstrate the efficacy of our proposed network in accurately classifying neonatal quiet sleep using EEG signals, which holds promising implications for clinical applications. By providing an automated and computationally efficient solution, our model can aid healthcare professionals in assessing neonatal sleep patterns and monitoring early-stage development. Building upon these findings, future work aims to develop an end-to-end deep learning model capable of classifying all three sleep stages using raw single-channel EEG signals. This advancement would eliminate the need for feature extraction and further streamline the classification process, enhancing the model's practicality and efficiency in real-time clinical applications. In conclusion, our research contributes to the field of neonatal sleep analysis by presenting a robust CNN-based approach for QS detection. The achieved accuracy, combined with the potential for further advancements, paves the way for

improved monitoring of neonatal sleep and early-stage development. Our work serves as a foundation for future studies aiming to refine and expand upon the capabilities of automated sleep stage classification in neonates.

Use of AI tools declaration

The authors declare they have not used artificial intelligence (AI) tools in the creation of this article.

Acknowledgments

This research work is funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R40), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Also, the authors would like to thank the neurologists at Fudan Children's Hospital for their kind support during data collection and annotation.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. A. R. Hassan, M. I. H. Bhuiyan, An automated method for sleep staging from EEG signals using normal inverse Gaussian parameters and adaptive boosting, *Neurocomputing*, **219** (2017), 76–87. https://doi.org/10.1016/j.neucom.2016.09.011
2. A. R. Hassan, M. I. H. Bhuiyan, A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features, *J. Neurosci. Methods*, **271** (2016), 107–118. https://doi.org/10.1016/j.jneumeth.2016.07.012
3. A. Dereymaeker, K. Pillay, J. Vervisch, S. Van Huffel, G. Naulaers, K. Jansen, et al., An automated quiet sleep detection approach in preterm infants as a gateway to assess brain maturation, *Int. J. Neural Syst.*, **27** (2017), 1750023. https://doi.org/10.1142/S012906571750023X
4. N. Koolen, L. Oberdorfer, Z. Rona, V. Giordano, T. Werther, K. Klebermass-Schrehof, et al., Automated classification of neonatal sleep states using EEG, *Clin. Neurophysiol.*, **128** (2017), 1100–1108. https://doi.org/10.1016/j.clinph.2017.02.025
5. K. Pillay, A. Dereymaeker, K. Jansen, G. Naulaers, S. Van Huffel, M. De Vos, Automated EEG sleep staging in the term-age baby using a generative modelling approach, *J. Neural Eng.*, **15** (2018), 036004. https://doi.org/10.1088/1741-2552/aaab73
6. J. Bronzino, *Principles of Electroencephalography*, CRC Press, 2015.
7. M. Andre, Pesquisas sobre formaço de professores: Contribuiçes delimitaço do campo, Convergncias e tenses no campo da formao e do trabalho docente: Didática, formaço de professores, trabalho docente, *Tech. Rep.*, 2010.
8. A. Loomis, E. Harvey, G. Hobart, Cerebral states during sleep, as studied by human brain potentials, *J. Exp. Psychol.*, **21** (1937), 127. https://doi.org/10.1037/h0057431

9. T. Lajnef, S. Chaibi, P. Ruby, P. E. Aguera, J. B. Eichenlaub, M. Samet, et al., Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines, *J. Neurosci. Methods*, **250** (2015), 94–105. https://doi.org/10.1016/j.jneumeth.2015.01.022
10. M. Xiao, H. Yan, J. Song, Y. Yang, X. Yang, Sleep stages classification based on heart rate variability and random forest, *Biomed. Signal Process. Control*, **8** (2013), 624–633. https://doi.org/10.1016/j.bspc.2013.06.001
11. P. Fonseca, N. den Teuling, X. Long, R. M. Aarts, Cardiorespiratory sleep stage detection using conditional random fields, *IEEE J. Biomed. Health Inf.*, **21** (2017), 956–966. https://doi.org/10.1109/JBHI.2016.2550104
12. S. Gudmundsson, T. P. Runarsson, S. Sigurdsson, Automatic sleep staging using support vector machines with posterior probability estimates, in *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, IEEE, (2005), 366–372.
13. H. Dong, A. Supratak, W. Pan, C. Wu, P. M. Matthews, Y. Guo, et al., Mixed neural network approach for temporal sleep stage classification, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **26** (2017), 324–333. https://doi.org/10.1109/TNSRE.2017.2733220
14. J. Zhang, R. Yao, W. Ge, J. Gao, Orthogonal convolutional neural networks for automatic sleep stage classification based on single-channel EEG, *Comput. Methods Programs Biomed.*, **183** (2020), 105089. https://doi.org/10.1016/j.cmpb.2019.105089
15. F. Andreotti, H. Phan, N. Cooray, C. Lo, M. T. Hu, M. De Vos, Multichannel sleep stage classification and transfer learning using convolutional neural networks, in *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, (2018), 171–174. https://doi.org/10.1109/EMBC.2018.8512214
16. J. Zhang, Y. Wu, A new method for automatic sleep stage classification, *IEEE Trans. Biomed. Circuits Syst.*, **11** (2017), 1097–1110. https://doi.org/10.1109/TBCAS.2017.2719631
17. A. Sors, S. Bonnet, S. Mirek, L. Vercueil, J. F. Payen, A convolutional neural network for sleep stage scoring from raw single-channel EEG, *Biomed. Signal Process. Control*, **42** (2018), 107–114. https://doi.org/10.1016/j.bspc.2017.12.001
18. T. F. Anders, R. N. Emde, A. H. Parmelee, *A Manual of Standardized Terminology, Techniques and Criteria for Scoring of States of Sleep and Wakefulness in Newborn Infants*, UCLA Brain Information Service/BRI Publications Office, NINDS Neurological Information Network, 1971.
19. J. W. Britton, L. C. Frey, J. L. Hopp, P. Korb, M. Z. Koubeissi, W. E. Lievens, et al., Electroencephalography (EEG): An introductory text and atlas of normal and abnormal findings in adults, children, and infants, *Am. Epilepsy Soc.*, (2016), 20–41.
20. R. J. Ellingson, Development of sleep spindle bursts during the first year of life, *Sleep*, **5** (1982), 39–46. https://doi.org/10.1093/sleep/5.1.39
21. J. Werth, L. Atallah, P. Andriessen, X. Long, E. Zwartkruis-Pelgrim, R. M. Aarts, Unobtrusive sleep state measurements in preterm infants–A review, *Sleep Med. Rev.*, **32** (2017), 109–122. https://doi.org/10.1016/j.smrv.2016.03.005
22. B. Chakravarthi, S. C. Ng, M. R. Ezilarasan, M. F. Leung, EEG-based emotion recognition using hybrid CNN and LSTM classification, *Front. Comput. Neurosci.*, **16** (2022), 1019776. https://doi.org/10.3389/fncom.2022.1019776

23. J. P. Turnbull, K. A. Loparo, M. W. Johnson, M. S. Scher, Automated detection of tracé alternant during sleep in healthy full-term neonates using discrete wavelet transform, *Clin. Neurophysiol.*, **112** (2001), 1893–1900. https://doi.org/10.1016/S1388-2457(01)00641-1
24. A. Piryatinska, G. Terdik, W. A. Woyczynski, K. A. Loparo, M. S. Scher, A. Zlotnik, Automated detection of neonate EEG sleep stages, *Comput. Methods Programs Biomed.*, **95** (2009), 31–46. https://doi.org/10.1016/j.cmpb.2009.01.006
25. A. H. Ansari, O. De. Wel, M. Lavanga, A. Caicedo, A. Dereymaeker, K. Jansen, et al., Quiet sleep detection in preterm infants using deep convolutional neural networks, *J. Neural Eng.*, **15** (2018), 066006. https://doi.org/10.1088/1741-2552/aadc1f
26. O. De Wel, M. Lavanga, A. Caicedo, K. Jansen, G. Naulaers, S. Van Huffel, Decomposition of a multiscale entropy tensor for sleep stage identification in preterm infants, *Entropy*, **21** (2019), 936. https://doi.org/10.3390/e21100936
27. L. Fraiwan, K. Lweesy, Neonatal sleep state identification using deep learning autoencoders, in *IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA)*, IEEE, (2017), 228–231. https://doi.org/10.1109/CSPA.2017.8064956
28. L. Fraiwan, M. Alkhodari, Neonatal sleep stage identification using long short-term memory learning system, *Med. Biol. Eng. Comput.*, **58** (2020), 1383–1391. https://doi.org/10.1007/s11517-020-02169-x
29. S. F. Abbasi, J. Ahmad, A. Tahir, M. Awais, C. Chen, M. Irfan, et al., EEG-Based neonatal sleep-wake classification using multilayer perceptron neural network, *IEEE Access*, **8** (2020), 183025–183034. https://doi.org/10.1109/ACCESS.2020.3028182
30. M. Awais, C. Chen, X. Long, B. Yin, A. Nawaz, S. F. Abbasi, et al., Novel framework: face feature selection algorithm for neonatal facial and related attributes recognition, *IEEE Access*, **8** (2020), 59100–59113. https://doi.org/10.1109/ACCESS.2020.2982865
31. M. Awais, X. Long, B. Yin, C. Chen, S. Akbarzadeh, S. F. Abbasi, et al., Can pre-trained convolutional neural networks be directly used as a feature extractor for video-based neonatal sleep and wake classification, *BMC Res. Notes*, **13** (2020), 1–6. https://doi.org/10.1186/s13104-020-05343-4
32. M. Awais, X. Long, B. Yin, S. F. Abbasi, S. Akbarzadeh, C. Lu, et al., A hybrid DCNN-SVM model for classifying neonatal sleep and wake states based on facial expressions in video, *IEEE J. Biomed. Health. Inf.*, **25** (2021), 1441–1449. https://doi.org/10.1109/JBHI.2021.3073632
33. S. F. Abbasi, M. Awais, X. Zhao, W. Chen, Automatic denoising and artifact removal from neonatal EEG, in *the Third International Conference on Biological Information and Biomedical Engineering*, VDE, (2019), 1–5.
34. S. F. Abbasi, H. Jamil, W. Chen, EEG-based neonatal sleep stage classification using ensemble learning, *CMC-Comput. Mater. Continua*, **70** (2022), 4619–4633. https://doi.org/10.32604/cmc.2022.020318
35. P. J. Cherian, R. M. Swarte, G. H. Visser, Technical standards for recording and interpretation of neonatal electroencephalogram in clinical practice, *Ann. Indian Acad. Neurol.*, **12** (2009), 58.
36. S. Janjarasjitt, M. S. Scher, K. A. Loparo, Nonlinear dynamical analysis of the neonatal EEG time series: the relationship between sleep state and complexity, *Clin. Neurophysiol.*, **119** (2008), 1812–1823. https://doi.org/10.1016/j.clinph.2008.03.024

37. D. Zhang, H. Ding, Y. Liu, C. Zhou, H. Ding, D. Ye, Neurodevelopment in newborns: a sample entropy analysis of electroencephalogram, *Physiol. Meas.*, **30** (2009), 491. https://doi.org/10.1088/0967-3334/30/5/006

AIMS Press