

Received 6 December 2023, accepted 14 December 2023, date of publication 22 December 2023,
date of current version 3 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3346059

RESEARCH ARTICLE

An Ensemble Voting Approach With Innovative Multi-Domain Feature Fusion for Neonatal Sleep Stratification

MUHAMMAD IRFAN^{1,2}, (Member, IEEE), HAFZA AYESHA SIDDIQA¹,
ABDELWAHEB NAHLIIS³, CHEN CHEN⁴, YAN XU⁵, LAISHUAN WANG⁶, ANUM NAWAZ^{1,2},
ABDULHAMIT SUBASI^{7,8}, TOMI WESTERLUND², (Senior Member, IEEE),
AND WEI CHEN^{1,4}, (Senior Member, IEEE)

¹Center for Intelligent Medical Electronics (CIME), School of Information Science and Technology, Fudan University, Shanghai 200437, China

²TIERS Group, Department of Computing, University of Turku, 20014 Turku, Finland

³Department of Circuits and Systems, School of Information Science and Technology, Fudan University, Shanghai 200437, China

⁴Human Phenome Institute, Fudan University, Shanghai 200437, China

⁵National Children's Medical Center, Children's Hospital of Fudan University, Shanghai 201102, China

⁶Children's Hospital of Fudan University, Shanghai 201102, China

⁷Institute of Biomedicine, Faculty of Medicine, University of Turku, 20520 Turku, Finland

⁸Department of Computer Science, College of Engineering, Effat University, Jeddah 21478, Saudi Arabia

Corresponding author: Wei Chen (w_chen@fudan.edu.cn)

This work was supported in part by the Shanghai Municipal Science and Technology International Research and Development Collaboration Project under Grant 20510710500, and in part by the Finnish National Agency for Education under Grant OPH-3323-2023.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Committee of the Children's Hospital affiliated to Fudan University under Approval No. (2017)89.

ABSTRACT A limited number of electroencephalography (EEG) channels are useful for neonatal sleep classification, particularly in the Internet of Medical Things (IoMT) field, where compact and lightweight devices are essential to monitoring health effectively. A streamlined and cost-effective IoMT solution can be achieved by utilizing fewer EEG channels, thereby reducing data transmission and device processing requirements. Using only two channels of an EEG device, this study presents a binary and multistage classification of neonatal sleep. The binary classification (sleep vs awake) achieved an accuracy of 87.56%, and a Cohen's kappa of 74.13%. The quiet sleep (Q_S) detection accuracy was 95.63%, with a Cohen's kappa of 83.87%. For the three-stage classification, accuracy was 83.72%, and Cohen's kappa was 69.73%. With only two channels, these are the highest performance parameters. The focus is on the fusion of features extracted through flexible analytical wavelet transform (FAWT) & discrete wavelet transform (DWT), ensemble-based voting models, and fewer channels. To feed crucial features into the ensemble-based voting model, feature importance, feature selection, and validation mechanisms were used. To design the voting classifier, several machine learning models were used, compared, and optimized. With SelectKBest feature selection, the proposed methodology was found to be the most effective. By using only two channels, this study shows the practicality of classifying neonatal sleep stages.

INDEX TERMS Dual channels, ensemble voting model, neonate sleep staging, EEG, internet of medical things, FAWT, DWT.

I. INTRODUCTION

The Internet of Medical Things (IoMT) offers the potential to improve neonatal health outcomes effectively and affordably

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Kafiul Islam¹.

through remote monitoring and diagnosis [1]. Using IoMT for health monitoring can improve accuracy, reliability, and convenience while reducing hospitalization and visitation costs [2], [3]. Additionally, this approach enhances data collection for research, resulting in a better understanding of neonatal sleep patterns and disorders [4]. It is important,

however, to optimize multiple domains when monitoring neonatal sleep patterns using Electroencephalography (EEG) devices to fully realize the potential benefits of IoMT devices. A high-quality EEG device with fewer channels can reduce physical and financial burdens in neonates due to their physiology and size. Because IoMT devices are small and low-power, they are well suited for neonatal sleep monitoring in remote areas where multi-channel and costly devices are not easily available. They can be developed more efficiently and cost-effectively by using fewer EEG channels. Further, by reducing the amount of data to be processed and transmitted to the cloud, the solution becomes more cost-effective [3], [5].

The proper assessment of neonatal sleep patterns is crucial for the early detection of health issues like sleep apnea and neurological disorders. EEG devices provide detailed information on different sleep stages, which play a vital role in healthy brain development and physical growth [6]. By categorizing EEG signals, abnormal sleep patterns can be detected accurately, allowing for timely interventions to improve neonatal health outcomes. Additionally, understanding a neonate's sleep staging can aid parents and caregivers in predicting when an infant will be awake or sleepy. Negligence in this area can lead to serious sleep-related issues like sudden infant death syndrome (SIDS) and accidental suffocation, and strangulation in bed (ASSB). Overall, research on neonatal sleep staging can provide valuable insights into early brain and body development [7], [8], [9].

Researchers are employing advanced signal processing techniques, machine learning (ML), and deep learning (DL) algorithms [10], [11], [12], [13]. These methods can help in extracting relevant features from the EEG signals, removing artifacts, and learning patterns that can distinguish between different sleep stages. Moreover, the development of standardized protocols for recording and analyzing neonatal EEG signals can also contribute to improving the accuracy and reliability of sleep stage classification.

However, neonatal EEG signals have different characteristics compared to adult EEG signals [14], [15]. Neonatal EEG signals typically exhibit lower frequency content, higher amplitude, and slow wave activity, making them challenging to classify accurately for sleep monitoring [16], [17]. Secondly, the brain development of neonates is rapid and dynamic, which results in changes in EEG patterns over a short period of time. This makes it difficult to establish consistent and reliable features for classification. Thirdly, the presence of various artifacts [18], such as muscle movements, eye blinks, and equipment noise, further complicates the analysis and interpretation of neonatal EEG signals. These artifacts can interfere with the underlying EEG patterns and lead to incorrect classification. Fourthly, sleep stages in neonates are not as well defined as in adults. Neonatal sleep is often divided into active sleep (A_S), quiet sleep (Q_S), and awake. However, these stages are not as clearly distinguishable as in adults, and transitions between stages

may be more subtle, making it harder to identify and classify sleep stages accurately. Fifthly, the lack of standardized recording and analysis techniques for neonatal EEG signals is another factor that complicates classification. There is no universally agreed-upon method for recording or processing neonatal EEG, leading to variations in data quality and interpretability [19].

Keeping in mind all of the obstacles in neonatal research for sleep monitoring, our study introduces a multidomain approach to analyzing neonatal EEG data, surpassing existing methods. This advancement is primarily due to our comprehensive feature extraction strategy, which integrates discrete wavelet transform (DWT), flexible analytical wavelet transform (FAWT), and features from frequency bands and temporal characteristics. This method is specifically tailored to address the unique properties of neonatal EEG data, which differ markedly from adult EEG signals. To determine the optimal combination for classifying neonatal sleep stages, we also assessed the effectiveness of several classifiers based on different ML techniques:

1. Extra Trees Classifier with Grid Search CV (ETCG)
2. Quadratic Discriminant Analysis (QDA)
3. Ensemble Random Forest Classifier (ERFC)
4. K Neighbors Classifier (KNN)
5. Artificial Neural Network (ANN)
6. Bagging Ensemble Model (BEM)
7. Extra Trees Classifier (ETC)
8. Gaussian NB (GNB)
9. Voting Classifier with Estimators:
 - Logistic Regression (LR)
 - Decision Tree Classifier
 - Support Vector Classifier
10. Voting Classifier with Estimators:
 - LR
 - GNB
 - Random Forest Classifier (RFC)
11. Stacking Classifier with Estimators:
 - KNN
 - Multi-layer Perceptron (MLP)
 - RFC
12. Gradient Boosting Classifier (GBC)
13. Ada Boost Classifier (ABC)
14. Multi-layer Perceptron (MLP)
15. Linear Discriminant Analysis (LDA)
16. Support Vector Classifier (SVM)

Main Contributions of Our Work Are as Follows:

Access to basic medical services is crucial for neonatal health, and the lack of such services can lead to preventable illnesses and complications. Addressing health facility disparities is important to improve neonatal outcomes, regardless of geographic location. The article's main contributions are:

- 1) This study has utilized a thorough multi-domain feature extraction methodology that includes DWT, FAWT, frequency band features, and temporal features. The FAWT along with the DWT framework, is a key

component of this methodology. FAWT is adept at breaking down signals into low-pass (LP) and high-pass (HP) channels, which is useful for adjusting important model parameters. The FAWT framework's capability to delve deeply into the signal characteristics makes it a vital tool in the analysis of non-stationary signals. This diverse feature extraction fusion improves the depth and breadth of signal analysis and serves as a solid foundation for the research.

- 2) To stratify neonatal sleep states, we developed an ensemble voting classifier using five base models, which were selected based on their superior performance across ten EEG channels as evidenced in experimental results. The algorithm was validated using datasets from a variety of participants and showed good classification accuracy for neonatal sleep states.
- 3) This study identifies neonatal sleep classification's two most effective EEG channels. Combining two channels further improves classification accuracy. The proposed model allows neonatal sleep monitoring with single and dual-channel settings since not all intensive care units have multi-channel EEG devices, and it could be very effective for remote areas.
- 4) The proposed stratification scheme efficacy is established across both binary and multiclass scenarios, employing the dataset outlined in [20]. To ensure robustness and reliability, a k-fold cross-validation methodology is used.

The paper is structured as follows: Section II reviews the relevant literature, while Section III elaborates on multi-domain feature extraction, feature selection, normalization, and employment of classifiers. Section IV provides the ablation study, illustrating how the proposed method is designed based on experimental results. The results, discussion, challenges, and comparisons to existing work are presented in Section V. Finally, the paper concludes in Section VI by summarizing the study's observations and outlining future research directions.

II. RELATED WORK

ML-based neonatal sleep staging automates the classification of sleep stages from EEG recordings, providing a more objective and reliable method than manual scoring by experts. ML models can quickly analyze large volumes of EEG data and detect subtle changes in sleep patterns that may be missed by human experts. This leads to increased consistency in sleep stage classification and improved diagnosis and treatment of underlying neurological or developmental issues in neonates [21]. Using large datasets and correctly annotated EEG recordings from neonates in various sleep states, ML and DL algorithms can be trained to become more accurate and generalizable.

Awais et al. achieved high accuracy in neonatal sleep classification for sleep and wake using a diffusion convolutional neural networks (DCNN) model for feature extraction and SVM for classification, with an accuracy of $93.8 \pm$

2.2% and an F1-score of 0.93 ± 0.3 [10]. However, their use of video EEG data raises privacy concerns due to the identifiable information it contains, such as neonates' faces and voices [22], [23]. In contrast, another study utilized an optimized Sinc-based DL model for classifying (Q_S) based on EEG data alone [24]. This approach is advantageous because neonates in the AS stage have EEG signals with a voltage amplitude similar to those in the awake stage, making it difficult to distinguish between them. However, DL models are typically more complex and require more parameters than ensemble ML models, resulting in higher computational demands [25]. Therefore, their applicability in the IoMT field may be limited due to high computational requirements.

Ansari et al. developed an 18-layer convolutional neural network (CNN) for automatic neonatal sleep state classification by separating (Q_S) from *non* - (Q_S) [26]. They used multi-channel EEG recordings of 26 preterm neonates and later introduced a multi-scale deep convolutional neural network for the same purpose [24]. By using a novel Sinc block, they were able to extract temporal features across multiple timescales.

Yu et al. performed automatic neonatal sleep state classification using publicly available single-channel EEG datasets [27]. They classified neonate's sleep patterns into W, N1, N2, N3, and REM based on the MRASleepNet module, which comprised a feature extraction module, a multi-resolution analysis (MRA) module, and a gated multi-layer perceptron (gMLP) module. Abbas et al. utilized single-channel EEG data for neonatal sleep-wake classification, employing a SVM algorithm. The results demonstrated an accuracy of 77.5% in sleep-wake classification, with a mean kappa of 55% [28].

Fraivan and Alkhodari used the long short-term memory (LSTM) learning technique to perform neonatal three-stage sleep classification by using EEG recordings of 16 full-term neonates [11]. Zhu et al. designed a novel multi-scale hierarchical neural network (MS-HNN) for the automatic classification of neonates' sleep with one, two, and eight channels [29]. They incorporated multi-scale convolutional neural networks (MSCNN), squeeze-and-excitation (SE) blocks, and temporal information learning (TIL) to extract more features that involve temporal information in sleep signals. With single and eight EEG channels, they achieved around 76.5% accuracy for three-stage classification.

In 2020, Abbasi et al. designed a deep MLP for (sleep vs awake) neonate sleep using multi-channel EEG recordings from 19 neonates [30]. They extracted a total of 12 features, including four frequency domains and eight time-domain features, from nine channels. Later, in 2021, they performed three-state neonatal sleep classification using the same EEG dataset [20] and CNN, SVM, and MLP classifiers, two ensemble algorithms for bagging and stacking were used. The study demonstrated that by using 9 channels, the classification accuracy for (sleep vs awake) was 82.53% while for Q_S v A_S v awake stage, it was around 81.99%. A single channel,

however, reduced the test accuracy to 71.7% for binary and 64.72% for three stages. The researchers also presented that by increasing the number of channels to four, classification accuracy improved to 73.15%. Another point to notice is that during the post-processing of EEG data, they applied a smoothing filter that halted the signal for a few minutes. This means the system cannot be used for real-time applications, especially for IoMT devices.

However, in this study, we have used only two channels, proposing a simplified data acquisition process for an EEG device, resulting in a reduction in data transfer to the cloud and making the solutions more cost-effective as discussed in our prior studies [3], [5].

III. MATERIALS AND METHODS

A. STUDY FLOWCHART

The flowchart depicted in Fig. 1 provides a simplified overview of the study methodology, illustrating the progression from the initial level to the final stage.

B. DATASET

Nineteen neonates were included in a study conducted at the NICU of the Children's Hospital of Fudan University, Shanghai, China. The study was approved by the Research Ethics Committee of the Children's Hospital (approval No. (2017) 89). The neonates underwent Video EEG (VEEG) recordings, and an average of 120 minutes of data was collected for each neonate, during which at least one sleep cycle was observed. A full 10-20 system for electrode placement, which included 17 electrodes, was used to acquire the EEG recordings. However, in this study, we have utilized 10 channels: F3-C3, C3-P3, F4-C4, C4-P4, F3-T3, T3-P3, F4-T4, T4-P4, T3-C3, C4-T4. Out of the 10 channels we tested, we selected the two that performed the best and analyzed their results. Further details of the dataset and data annotation are provided in previous studies [10], [20], [30].

C. PREPARATION AND PRE-PROCESSING OF DATA

EEG may contaminate with noise and artifacts, which could compromise its accuracy and reliability. EEG recordings were originally sampled at 500 Hz, and unwanted signals needed to be removed before further processing. A three-step pre-processing method was employed to remove noise and artifacts from EEG recordings. Firstly, we used a Finite Impulse Response (FIR) filter (high-pass (HP) = 0.3 Hz and low-pass (LP) = 35 Hz) from EEGLAB to filter the EEG signals. EEGLAB offers more advanced EEG filtering capabilities than other packages [31]. During this step, the unwanted frequency range was removed. Second, we segmented the multichannel EEG into epochs of 30 seconds and labeled them accordingly. Segmentation made it easier to analyze the EEG data since it divided it into manageable segments. In the final step, we removed artifacts that constituted around 20% of the EEG data post-segmentation, and only noise-free recordings were used to

TABLE 1. The frequencies related to the various levels of decomposition for a db-4 filter wavelet with a 500 Hz sampling frequency.

Decomposed signal	Frequency range [Hz]
D1	125–250
D2	62.5–125
D3	31.25–62.5
D4	15.625–31.25
D5	7.8125–15.625
D6	3.90625–7.8125
D7	1.953125–3.90625
A7	0–1.953125

train and test the model. To validate the result, 7-fold cross-validation is used.

D. MULTI-DOMAIN FEATURE EXTRACTION

Neonatal EEG signals vary distinctly from adult signals due to intra-class differences based on brain structure and function. Automatic sleep stage detection in the neonatal domain faces challenges such as limited datasets [32], vague data descriptions, and low-precision detection methods. Hence, it is crucial to improve feature extraction techniques to enable accurate and precise sleep stage detection. A total of 173 features were extracted from each epoch: 12 from the spectral statistics of four bands (α , β , θ , δ), 11 from temporal statistics, 55 from DWT, and 95 from FAWT.

1) DISCRETE WAVELET TRANSFORM

The DWT is a commonly used technique for the time-frequency analysis of signals [33], [34]. It decomposes a signal into a linear combination of translations and dilations of a basis function called the ‘‘Mother Wavelet,’’ as well as a scaling function. During the DWT process, the signal is filtered using an LP filter and a HP filter, resulting in approximate coefficients A_i and detailed coefficients D_i . These coefficients are then downsampled, and the process is repeated to obtain the next level of approximate and detailed coefficients. This iterative process establishes time-scale regions, rather than time-frequency regions, for the signal.

In this study, we concentrated on EEG signals, which generally lack useful frequencies above 30 Hz. As a result, we opted for 7 decomposition levels enabling us to maintain the signal segments that strongly correlate with the frequencies necessary for classification. This led to the EEG signals being decomposed into seven detail coefficients (D_1 - D_7) and one final approximation coefficient (A_7). Furthermore, we chose the Daubechies wavelet of order 4 (db4) for detecting changes in EEG signals as it offers a higher accuracy score at the cost of computation.

Features Derived from DWT:

The wavelet coefficients obtained offer a concise representation that illustrates the energy distribution of the EEG signal in both time and frequency domains. Table 1 displays the frequencies corresponding to various decomposition levels for the db4 wavelet, given a sampling frequency of 500 Hz.

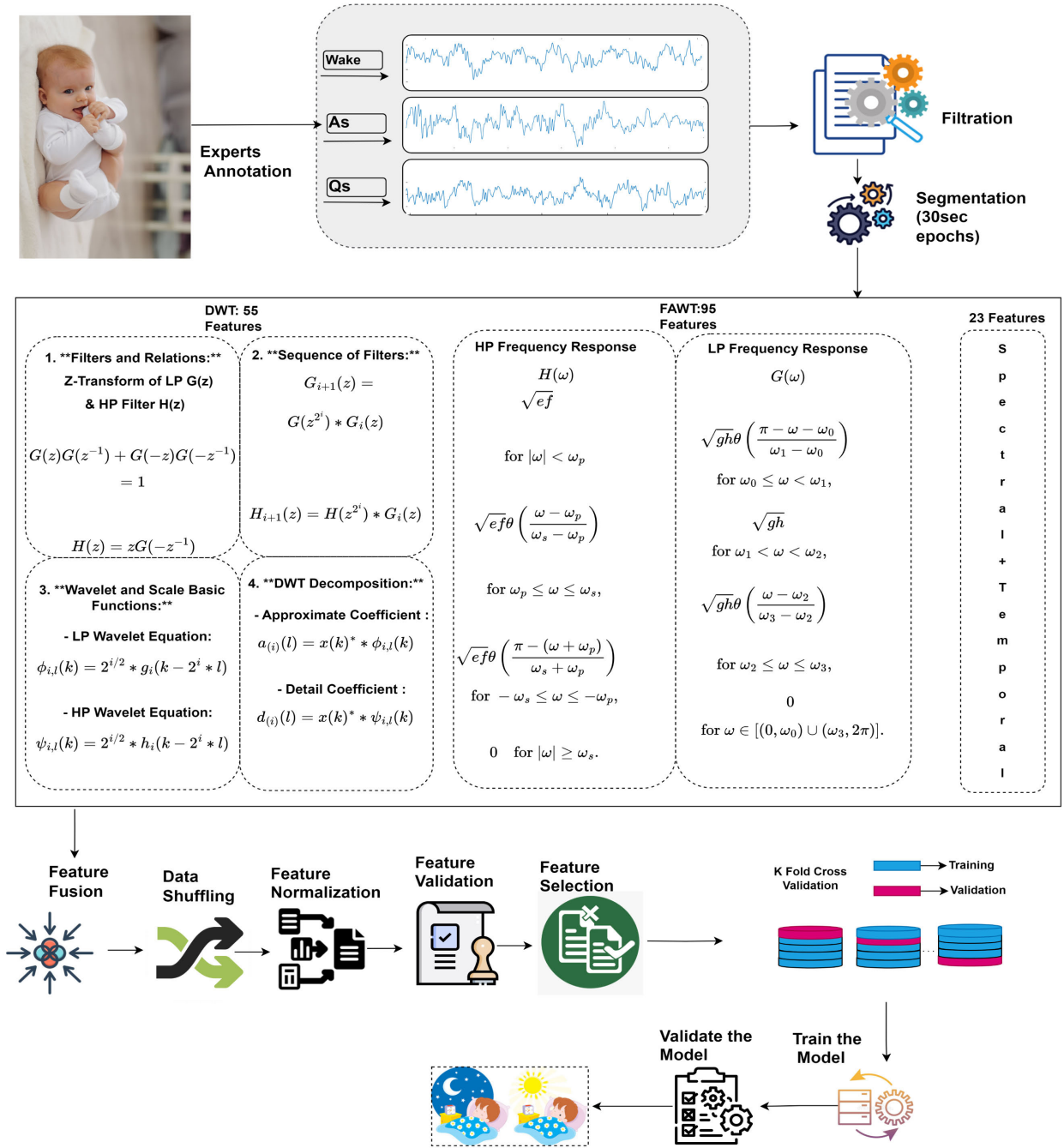


FIGURE 1. Flowchart and methodological overview. (The neonate picture is from freepik.com and we have a premium plan).

To represent the time-frequency distribution of EEG signals, the study used the following statistical features:

- 1) The mean of the absolute values of coefficients in each sub-band;
- 2) The median of coefficients in each sub-band;
- 3) The root mean square values of each sub-band;
- 4) The standard deviation of coefficients in each sub-band;
- 5) The ratio of absolute mean values of adjacent sub-bands;
- 6) The skewness of each sub-band;
- 7) The kurtosis of each sub-band.

2) FLEXIBLE ANALYTIC WAVELET TRANSFORM

A more comprehensive view of time-frequency attributes of signals is offered by FAWT, as described in [35]. This method is particularly suitable for analyzing oscillatory signals since it includes Hilbert transform atom pairs. FAWT is governed by three parameters: the Quality Factor Q , the Number of Decompositions J , and the Redundancy Factor r . The objective of Q is to control oscillations in the mother wavelet, which is defined as a function of frequency variables and a constant β as

follows:

$$Q = \frac{\omega_0}{\Delta\omega}, \quad \beta = \frac{2}{Q+1} \quad (1)$$

where, ω_0 and $\Delta\omega$ represent the central frequency and bandwidth of the signal, respectively. The redundancy factor r deals with the time-localization aspect of the wavelet. An iterative filter bank of HP and LP channels is employed in FAWT. This configuration allows custom tuning of the dilation factor, Q , and r through parameter adjustments in constants β and e, f, g, h . The parameters e and f are set for up and down-sampling of the HP filter, and g and h are configured for up and down-sampling of the LP channel. The process computes J decomposition levels in an iterative manner, and each level is comprised of LP and HP channels that differentiate the negative and positive frequencies, respectively. The frequency responses $H(\omega)$ and $G(\omega)$ of the HP and LP filters are as follows:

$$H(\omega) = \begin{cases} \sqrt{ef} & \text{for } |\omega| < \omega_p, \\ \sqrt{ef}\theta \left(\frac{\omega - \omega_p}{\omega_s - \omega_p} \right) & \text{for } \omega_p \leq \omega \leq \omega_s, \\ \sqrt{ef}\theta \left(\frac{\pi - (\omega + \omega_p)}{\omega_s + \omega_p} \right) & \text{for } -\omega_s \leq \omega \leq -\omega_p \\ 0 & \text{for } |\omega| \geq \omega_s. \end{cases}$$

$$G(\omega) = \begin{cases} \sqrt{gh}\theta \left(\frac{\pi - \omega - \omega_0}{\omega_1 - \omega_0} \right) & \text{for } \omega_0 \leq \omega < \omega_1, \\ \sqrt{gh} & \text{for } \omega_1 < \omega < \omega_2, \\ \sqrt{gh}\theta \left(\frac{\omega - \omega_2}{\omega_3 - \omega_2} \right) & \text{for } \omega_2 \leq \omega \leq \omega_3, \\ 0 & \text{for } \omega \in [(0, \omega_0) \cap (\omega_3, 2\pi)]. \end{cases}$$

Parameters associated with the above-mentioned filter banks are as follows:

$$\omega_p = (1 - \beta)\pi + e^e, \quad (2)$$

$$\omega_s = \pi f, \quad (3)$$

$$\omega_0 = (1 - \beta)\pi + e^g, \quad (4)$$

$$\omega_1 = e\pi fg, \quad (5)$$

$$\omega_2 = \pi - e^g, \quad (6)$$

$$\omega_3 = \pi + e^g, \quad (7)$$

$$\epsilon \leq e^{-f+\beta fe+f\pi}. \quad (8)$$

where r indicates the ratio of input to output samples, which should be greater than one to avoid information loss. It is defined as:

$$r = \left(\frac{g}{h} \right) \frac{1}{1 - \frac{e}{f}} \quad (9)$$

For a precise reconstruction, β should be less than one, and it is defined as:

$$1 - \frac{e}{f} \leq \beta \leq \frac{g}{h} \quad (10)$$

The shift-invariance, tunable oscillatory conditions, and flexible time-frequency localization of FAWT make it useful for a variety of practical applications [36].

Features Derived from FAWT:

The FAWT implementation entails choosing parameters for seven decomposition levels. For dual-state EEG signals, the subband signals are subsequently reconstructed in descending frequency order. The e/f value is set at 3/4, and the values of r and Q are pre-specified as discussed in our previous study [35]. Furthermore, g/h is set at 1/2 to maintain constraints on r for information preservation, as dictated in (9). Since FAWT offers a high dimensional feature space for searching features, compact statistics such as mean, standard deviation, skewness, and kurtosis are evaluated for each signal subband. The use of such statistical measures is promising for capturing inherent signal information [37]. Six such features are listed below:

- 1) Mean absolute value of coefficients within each subband;
- 2) Average power within each subband;
- 3) Standard deviation within each subband;
- 4) Ratio of mean absolute values of adjacent subbands;
- 5) Skewness within each subband;
- 6) Kurtosis within each subband.

3) SPECTRAL STATISTICS OF FOUR BANDS

Different frequencies of waves in EEG signals, such as α , β , θ , and δ , are linked with different stages of sleep and cognitive and emotional states. The statistical features that we calculated for each EEG frequency band are the mean, median, and standard deviation. From the spectral statistics of four bands, in total, twelve features were calculated to analyze the relationship between EEG frequency bands and sleep states. These statistical features provide information about the amplitude and variability of each frequency range.

4) TEMPORAL FEATURES OF EEG SIGNALS

Time-domain statistical features are a critical tool for analyzing neonatal sleep patterns as derived from EEG signals. Specifically, eleven features were calculated, including mean amplitude, standard deviation, peak-to-peak amplitude, variance, square root, minimum amplitude, maximum amplitude, root mean square, and absolute mean change in amplitude, skewness, and kurtosis

These features have been shown to provide critical information about the underlying neural processes in an efficient and effective manner. Additionally, the analysis also incorporated spectral statistics of four bands and FAWT-DWT-derived features. Overall, the combined use of these features offers a promising approach for developing automated sleep staging schemes that could lead to improved diagnosis and treatment of neonatal sleep-related issues.

E. FEATURE NORMALIZATION

We processed the data in two ways for normalization to reduce overfitting. As a first step, normalize the data by

scaling each row to a unit norm. We used the L2 norm, or Euclidean norm, for this. L2 norms are calculated by in (11).

$$x_{normalized} = x/\sqrt{\sum(x^2)} \quad (11)$$

where x is a vector and $x_{normalized}$ the normalized vector.

Min-Max scaling is the next step. In this method, the features in the data are scaled to a given range between 0 and 1.

$$x_{scaled} = (x - \min(x))/(\max(x) - \min(x)) \quad (12)$$

where x is a vector, $\min(x)$ the minimum value of x , $\max(x)$ the maximum value of x , and x_{scaled} is the scaled vector.

By improving data consistency and limiting the impact of extreme values, these methods can help reduce overfitting. These can also improve model performance on data that has yet to be analyzed and promote more model generalization.

F. FEATURE IMPORTANCE AND SELECTION

The SelectKBest method was used to select the features. Feature selection is used to identify and select the most relevant features in a dataset and exclude the less relevant ones. The SelectKBest (SelectKBest(score_func=f_classif, k=max_features)) uses a scoring function, such as f_classif, to calculate the significance of each feature with respect to the target variable. The scoring function is based on the ANOVA F-value, which measures the difference between the means of two groups. The k parameter specifies how many top features should be selected. Lastly, only the selected features are returned from the transformed input data [38]. In high-dimensional datasets, feature selection is particularly important because it reduces the dimensionality of the dataset and can improve the performance of ML models.

G. CLASSIFIERS

In this study, we evaluated and compared several different classifiers using various ML techniques to design an optimal voting classifier. The voting classifier combines the predictions of multiple base classifiers to generate a more robust and generalized model. Essentially, it's like seeking a "second opinion" from other models to make a more informed decision. In this study, we have utilized hard voting. Each model in the ensemble "votes" for a class, and the class that receives the majority of votes is the chosen prediction. The ensemble approach outperforms the individual models in this study because it takes into account each classifier's strengths and mitigates its weaknesses.

Overview of Base Classifiers:

Each model's performance metrics on each channel determine the most appropriate base classifier. Additionally, the results of applying the models to the two most effective channels are considered in this decision-making process. The performance outcomes for the individual and combined channel applications are detailed in Table 2 and Table 3. The tables provide information about the effectiveness of each

model, making it easier to choose the best classifier based on data. It ensures that the choice is based on empirical data and highlights the advantages and disadvantages of each model in terms of channel-specific and integrated channel performance.

1) EXTRATREESCLASSIFIER

The Extra Trees classifier is an ensemble approach. Splits are selected randomly rather than based on best splits, so it introduces more randomness. A reduction in variance may be achieved by this randomness at the expense of a slight increase in bias.

$$H(T) = - \sum_i p_i \log p_i \quad (13)$$

where $H(T)$ represents the entropy of the target variable T , and p_i represents the proportion of the target variable's category i in the dataset.

2) DECISIONTREECLASSIFIER

In decision trees, data is split recursively based on attribute values, and subsets are measured by metrics like Gini impurity in order to achieve pure or nearly pure subsets.

$$Gini(T) = 1 - \sum_i p_i^2 \quad (14)$$

3) RANDOMFORESTCLASSIFIER

A Random Forest classifier is an ensemble of decision trees that are trained on different subsets of data and features. Predictions are averaged to improve generalization and robustness.

$$O_{RF} = \frac{1}{N} \sum_{i=1}^N O_i \quad (15)$$

4) MLPCLASSIFIER

MLP classifiers are artificial neural networks that have three layers: an input layer, a hidden layer, and an output layer. In particular, MLPs excel at capturing complex relationships in data. This study, however, used a single hidden layer with 1000 neurons.

5) BAGGINGCLASSIFIER

The Bagging (Bootstrap Aggregating) classifier is an ensemble model, which trains multiple instances of a base classifier (in our case, a Decision Tree), each on a random subset of training data.

H. HYBRID ENSEMBLE VOTING MODEL

This study presents a hybrid ensemble voting model to optimize the reliability and accuracy of neonatal sleep stage classification. It consists of a DecisionTreeClassifier, four configurations of ExtraTreesClassifier, RandomForestClassifier, MLPClassifier, and a BaggingClassifier with a DecisionTree base estimator. The hybrid ensemble approach

integrates the strengths of different base models, including decision trees for their interpretability, Extra Trees and Random Forests for their randomness and generalizability, and neural networks for capturing complex relationships in neonatal EEG data. A voting mechanism is used to aggregate predictions from these base models, thereby producing a final classification that integrates the strengths and mitigates the weaknesses of each model. Based on our extensive experimental evaluations, the approach is robust and accurate. The model pseudo code is provided in algorithm 1.

Algorithm 1 Ensemble Voting Classifier With Multiple Base Classifiers and K-Fold Cross-Validation

- 1: Define various base classifiers:
multiple **ExtraTreesClassifiers** (e.g., 540, 900, 470, 300 estimators),
a **DecisionTreeClassifier**,
a **RandomForestClassifier** (e.g., 400 estimators),
a **MLPClassifier** (e.g., single hidden layer with 1000 neurons),
a **BaggingClassifier** using **DecisionTree** as base estimator (e.g., 1000 estimators).
 - 2: Combine these classifiers into a **voting classifier**, voting=hard.
 - 3: Set the number of folds 7 for **k-fold cross-validation** and create a k-fold cross-validator **kf**.
 - 4: Initialize empty **lists** to store **evaluation metrics (accuracies, precisions, recalls, f1_scores, cohen_kappas)**.
 - 5: **for** each **fold** in the k-fold cross-validation **do**
 - 6: Split the **data** into **training** and **validation sets**.
 - 7: Train the **voting classifier** using the **training data**.
 - 8: Predict the **labels** for the **validation data**.
 - 9: Calculate **evaluation metrics** for the **current fold** and append them to the respective **lists**.
 - 10: Print the **metrics** for the **current fold**.
 - 11: **end for**
 - 12: Calculate the **average** of the **evaluation metrics** across **all folds**.
 - 13: Print the **average evaluation metrics**.
-

I. EVALUATION METRICS FOR CLASSIFICATION MODELS

We have used five key metrics to evaluate classification models, each providing insights into a different aspect of performance. In general, accuracy measures correctness, while precision and recall measure positive predictions' accuracy. To provide a balanced view, the F1 score harmonizes precision and recall. Additionally, Cohen's Kappa provides a robust measure of model accuracy that accounts for chance agreement. A classification model's assessment and improvement are guided by these metrics.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (16)$$

where TP (True Positive) and TN (True Negative) represent correct positive and negative predictions, respectively. FP

(False Positive) and FN (False Negative) signify incorrect positive and negative predictions, respectively.

$$Precision = \frac{TP}{(TP + FP)} \quad (17)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (18)$$

$$F1 - score = \frac{2 * (precision * recall)}{(precision + recall)} \quad (19)$$

$$k = \frac{(P_o - P_e)}{(1 - P_e)} \quad (20)$$

where k denotes Cohen's kappa, P_o observed agreement between raters (proportion of cases where the raters agreed), and P_e expected agreement between raters by chance based on the distribution of ratings.

IV. ABLATION STUDY

This ablation study presents a comprehensive analysis of sixteen ML models (names are given in section I), encompassing ensemble methods, stacking, and voting classifiers. The objective is to categorize neonatal sleeping stages accurately. We examined various feature selection techniques, focusing particularly on the SelectKBest method, within both binary and multiclass classification frameworks.

A. SINGLE CHANNEL ANALYSIS USING MULTIDOMAIN FEATURE

1) BINARY CLASSIFICATION: SLEEP VS AWAKE

The fusion of FAWT, DWT, spectral, temporal features (FDSTF) method outperformed the combination of DWT, spectral, temporal features (DSTF) method in the binary classification of neonatal EEG data (sleep vs awake), as shown in Fig. 2. In the EEG channel F3-T3, the ensemble voting model with FDSTF achieved an accuracy of 84.20%, surpassing DSTF's 80.69%. Similarly, in the C4-T4 channel, the ensemble voting model with FDSTF achieved 85.80% accuracy, outperforming DSTF's 82.21%. Nevertheless, Kappa scores for both methods showed a decline in classification consistency.

2) THREE-STAGE CLASSIFICATION: Q₅ VS A₅ VS AWAKE

FDSTF outperformed DSTF in the more complex three-stage classification, as depicted in Fig. 3. F3-T3 and C4-T4 achieved accuracies of 80.71% and 80.57%, respectively, for FDSTF, whereas DSTF achieved 80.28% and 77.66%, respectively. In terms of Kappa scores, DSTF scored 62.22% and 56.96%, while FDSTF achieved 63.91% and 63.83% for F3-T3 and C4-T4, respectively.

B. COMPARATIVE ANALYSIS OF FEATURE SELECTION TECHNIQUES

A detailed comparative analysis is presented, evaluating three feature selection techniques—Principal Component Analysis (PCA), SelectKBest, and SelectPercentile—for two classification tasks: (sleep vs awake) and (Q₅ vs A₅ vs

TABLE 2. Model performance for binary classification on two best channels.

Model	sleep vs awake (C4-T4)		sleep vs awake (F3-T3)		sleep vs awake (F3-T3 , C4-T4)	
	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
Extra Trees Classifier	84.98	70.10	83.71	66.09	86.07	73.84
Quadratic Discriminant Analysis	49.24	84.30	56.91	18.24	61.17	22.94
Random Forest Classifier	85.23	69.54	82.86	64.31	85.92	72.79
K Neighbors Classifier	76.42	51.79	68.84	36.46	-	-
BaggingClassifier	82.86	64.72	82.58	63.67	85.13	69.32
Gaussian NB	48.67	0.070	47.06	0.05	48.67	0.0695
Gradient Boosting Classifier	83.62	66.21	81.66	65.29	84.98	70.21
AdaBoost Classifier	80.40	59.68	80.02	59.13	81.82	62.78
MLP Classifier	81.06	61.09	76.93	57.50	85.13	69.71
LinearDiscriminantAnalysis	76.42	51.64	79.17	57.16	80.40	59.82
Support Vector Machine	73.48	44.31	78.03	54.12	81.16	61.08

TABLE 3. Models performance for three stage classification on two best channels.

Model	Q_S vs A_S vs awake (C4-T4)		Q_S vs A_S vs awake (F3-T3)		Q_S vs A_S vs awake (C4-T4, F3-T3)	
	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
Extra Trees Classifier	76.12	56.29	79.44	61.77	82.67	68.27
Quadratic Discriminant Analysis	54.17	26.67	49.34	19.33	51.89	22.43
Random Forest Classifier	75.18	55.37	79.17	61.05	80.78	64.42
KNeighbors Classifier	67.71	40.26	64.58	32.66	74.43	53.56
Bagging Classifier	75.61	55.37	77.46	57.92	79.36	62.00
Gaussian NB	28.50	0.09	26.04	0.069	48.67	0.0695
Gradient Boosting Classifier	75.12	54.80	78.79	60.18	80.04	63.49
AdaBoost Classifier	72.16	48.94	73.39	52.31	73.77	52.58
MLP Classifier	75.43	54.88	76.93	57.50	78.60	62.22
LinearDiscriminantAnalysis	67.23	38.65	71.50	47.73	74.24	53.57
Support Vector Machine	65.25	28.99	71.59	44.87	74.91	53.08

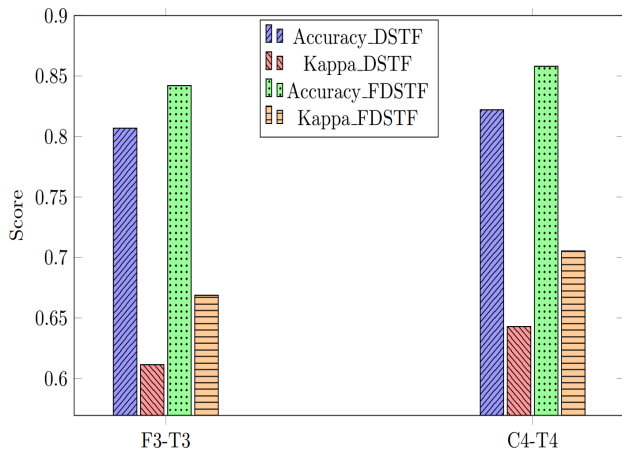


FIGURE 2. DSTF vs FDSTF: sleep vs awake.

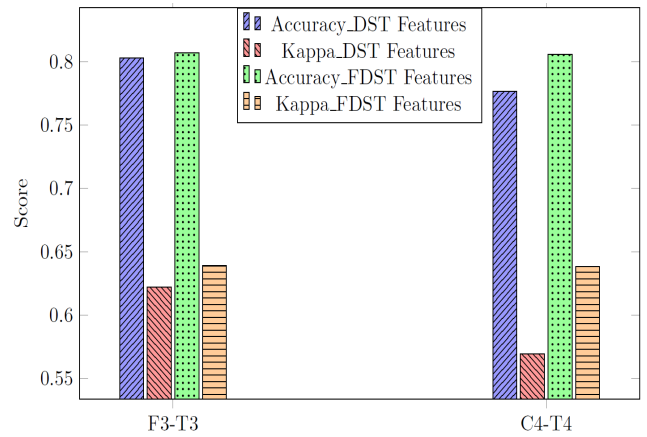


FIGURE 3. DSTF vs FDSTF: Q_S vs A_S vs awake.

awake). Fig. 4 elucidates the performance scores for each method within these contexts, thereby guiding the selection of the most effective feature selection technique.

In the (sleep vs awake) task, SelectKBest demonstrates superior performance, achieving a score of 0.8756, closely trailed by SelectPercentile with a score of 0.8741. PCA, while effective, scores a lower 0.8085. For the (Q_S vs A_S vs awake) classification, SelectKBest leads again, with a score of 0.8384, followed by SelectPercentile at 0.8367. PCA shows a more distinct performance gap, scoring 0.7358. These findings, as depicted in Fig. 4, are crucial for identifying the most suitable feature selection method for each task based on empirical performance data.

C. SELECTION OF BEST CHANNELS

This segment involved a thorough analysis of individual EEG channels in binary and multi-stage classifications using multi-domain features, feature selection, and ensemble modeling techniques. The evaluation focused on accuracy and Kappa value. In binary classification, the C4-T4 channel showed notable effectiveness, while in three-stage classification, the F3-T3 channel stood out, indicating its utility in more nuanced EEG applications.

D. SELECTION OF BEST ML MODELS

The performance of various ML models was evaluated in both binary (sleep vs awake) and three-stage (Q_S vs A_S

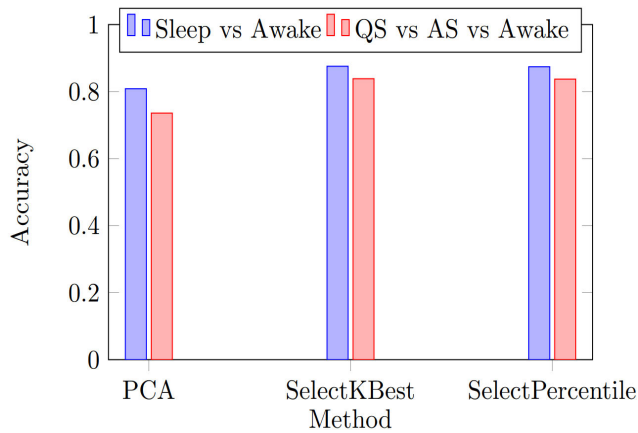


FIGURE 4. Comparison of feature selection techniques across different scenarios.

TABLE 4. Cross-validation results for each fold: sleep vs awake.

Fold	Accuracy	Precision	Recall	F1	Kappa
1	89.66	89.91	89.66	89.53	78.16
2	88.27	88.67	88.27	88.16	76.04
3	87.48	87.53	87.48	87.40	74.15
4	85.69	85.97	85.69	85.55	70.66
5	84.49	84.66	84.49	84.34	67.98
6	89.66	89.75	89.66	89.52	77.50
7	87.65	87.63	87.65	87.60	74.41
Mean	87.56	87.73	87.56	87.44	74.13

vs awake) classifications using channels C4-T4 and F3-T3, as well as dual-channel setup. Tables 2 and 3 detail these models' effectiveness in differentiating neonatal sleep stages.

V. RESULT AND DISCUSSION

A. CROSS-VALIDATION RESULTS FOR SLEEP VS AWAKE: DUAL CHANNEL (F3-T3, C4-T4 ANALYSIS)

In Table 4, we present the 7-fold cross-validation results for (sleep vs awake) classification using dual channels. An ensemble-based voting classifier using FDSTF was used to obtain the results. For each fold, the table shows performance metrics such as accuracy, precision, recall, F1 score, and the kappa statistic. All folds achieved an average accuracy of 87.56%, and the mean kappa value was 74.13%.

B. CROSS-VALIDATION RESULTS FOR Q_S DETECTION: DUAL CHANNEL (F3-T3, C4-T4 ANALYSIS)

For Q_S detection, we demonstrate the results of a 7-fold cross-validation in Table 5. The results were achieved using an ensemble-based voting classifier utilizing FDSTF. For each fold of the validation process, accuracy, precision, recall, F1 score, and Cohen's kappa statistic are reported. Overall, the proposed methodology achieved an impressive average accuracy of 95.63% and a mean kappa statistic of 83.87%, reflecting good agreement between predictions and actual labels.

TABLE 5. Cross-validation results for each fold: Q_S detection.

Fold	Accuracy	Precision	Recall	F1	Kappa
1	96.65	96.71	96.65	96.53	87.68
2	95.54	95.47	95.54	95.42	84.64
3	94.42	94.42	94.42	94.19	81.12
4	94.64	94.46	94.64	94.45	76.93
5	97.77	97.83	97.77	97.70	90.89
6	95.09	95.10	95.09	94.87	81.84
7	95.31	95.29	95.31	95.16	83.97
Mean	95.63	95.61	95.63	95.48	83.87

TABLE 6. Cross-validation results for each fold: (Q_S vs A_S vs Awake).

Fold	Accuracy	Precision	Recall	F1	Kappa
1	84.10	83.90	84.10	83.38	69.69
2	82.90	83.12	82.90	82.32	69.36
3	85.49	85.47	85.49	85.04	73.12
4	80.12	79.96	80.12	79.18	63.69
5	80.52	80.48	80.52	79.84	64.05
6	88.07	87.93	88.07	87.64	76.31
7	84.86	84.79	84.86	84.53	71.86
Mean	83.72	83.66	83.72	83.13	69.73

TABLE 7. Performance comparison of proposed method and existing work for sleep vs awake detection.

Study	Channels	Accuracy	Kappa
[30]	9	82.53	65.00
[30]	1	71.00	52.00
[30]	4	74.70	55.00
[20]	9	85.66	69.00
Proposed	2	87.56%	74.13

C. CROSS-VALIDATION RESULTS FOR Q_S VS A_S VS AWAKE STATE: DUAL CHANNEL (F3-T3, C4-T4 ANALYSIS)

Table 6 provides the results of a 7-fold cross-validation to distinguish between (Q_S vs A_S vs awake). An ensemble-based voting classifier with FDSTF was designed to achieve these outcomes. Each fold of the validation process includes metrics such as accuracy and Cohen's kappa. In the proposed methodology, an average accuracy of 83.72% was achieved and a kappa statistic of 69.73% was reached.

D. COMPARISON OF PROPOSED METHOD WITH EXISTING WORK

Table 7 shows that our methodology for (sleep vs awake), which utilizes data from two channels, an ensemble-based voting classifier, and FDSTF outperforms existing approaches with an accuracy of 87.56% and a Kappa value of 74.13%. Moreover, Abbasi et al. [20], [30], used nine channels and reached 82.53% accuracy score.

In Table 8, we show that our proposed Q_S detection method has a higher accuracy 95.63% as well as a high Kappa value (83.87%), which implies a higher level of reliability than existing models by Abbasi et al. [20], Moghadam et al. [18] and other studies mentioned in the Table 8.

For the three-stage classification of neonatal sleep, our method achieves 83.72% accuracy and 69.73% Kappa value,

TABLE 8. Performance comparison of proposed method and existing work for Q_S detection.

Study	Channels	Accuracy	Kappa
[20]	9	94.27	73.00
[18]	4	86.00	76.00
[24]	8	–	77.00
[24]	1	–	75.00
Proposed	2	95.63	83.87

TABLE 9. Performance comparison of proposed method and existing work for three stage classification (Q_S vs A_S vs awake).

Study	Channels	Accuracy	Kappa
[20]	9	81.99	66.00
[29]	8	76.50	73.50
[39]	–	80.40	–
[40]	–	82.50	–
[41]	–	80.66	66.70
Proposed	2	83.72	69.73

compared with other methods that utilize more channels as shown in Table 9. Overall, our proposed approach utilizes fewer channels and ML techniques while exhibiting superior or comparable performance.

E. DISCUSSION

Our research represents an advance in the field of neonatal sleep stratification, which has long struggled with the intricacies and irregularities present in neonatal data, which are frequently exacerbated by outside interferences like feeding and movement. Specifically, our work introduces novel methodologies designed for the analysis of neonatal EEG data, which distinguishes it from typical methods that are primarily focused on adult EEG data.

One of the main focuses of our research is dealing with the peculiarities of neonatal EEG signals, which are characterized by a higher amplitude and a lower frequency in comparison to adult EEG, which poses special analytical challenges. Our approach, which is carefully tailored to these particular characteristics, has proven to be successful in accurately classifying neonatal sleep states, which is especially important considering the rapidity of neonatal brain development and the resulting evolution of their EEG patterns.

The most significant of our contributions is the development of a novel, all-encompassing multi-domain feature extraction methodology. This methodology combines methods like DWT, FAWT, and spectral and temporal features. The use of the FAWT framework is very important since it can intricately break down non-stationary signals, which are commonly observed in newborn EEG, and allow for a thorough examination of signal complexities, enhancing our understanding and analytical precision. Another noteworthy aspect of our research is the development of an ensemble voting classifier and the optimized feature selection process. SelectKbest was found to be the most efficient

method for processing neonatal EEG data through extensive experimentation. We further refined our methodology by using an ensemble of the top five ML models, which were selected based on how well they performed across ten EEG channels. The base classifiers in this voting ensemble model are included based on how well they perform across different channels. The selection of base models was grounded in an extensive, empirical study of model performance across multiple EEG channels. This practical approach highlights the usefulness of our method in real-world neonatal care settings, where accuracy and precision are crucial, as well as its dependability.

In the future, while our current methodology adeptly addresses binary to three neonatal sleep stages, it is essential to recognize that neonates exhibit five distinct sleep stages. Consequently, future endeavors will aim to extend the applicability of our model to encompass all these stages. Moreover, our research trajectory includes exploring novel feature extraction methods, integrating with IoT devices for real-time monitoring, and employing advanced ML architectures like recurrent neural networks and transformers.

VI. CONCLUSION

In this research, we demonstrate that a fusion of features extracted from FAWT and DWT, along with temporal and spectral features, is highly reliable for classifying neonatal sleep stages. In particular, our ensemble-based voting model yielded an accuracy of 87.56% for (sleep vs awake), 95.63% for Q_S , and 83.72% for differentiating between (Q_S , A_S , and awake). Our approach requires only two EEG channels, making it a cost-effective and efficient solution for the Internet of Medical Things (IoMT). Through the implementation of a voting model and thorough validation techniques, we have developed an accurate and practical approach to neonatal sleep stage classification.

REFERENCES

- [1] B. Kim, S. Kim, M. Lee, H. Chang, E. Park, and T. Han, "Application of an Internet of Medical Things (IoMT) to communications in a hospital environment," *Appl. Sci.*, vol. 12, no. 23, p. 12042, Nov. 2022.
- [2] Z. Ashfaq, A. Rafay, R. Mumtaz, S. M. H. Zaidi, H. Saleem, S. A. R. Zaidi, S. Mumtaz, and A. Haque, "A review of enabling technologies for Internet of Medical Things (IoMT) ecosystem," *Ain Shams Eng. J.*, vol. 13, no. 4, Jun. 2022, Art. no. 101660.
- [3] M. Irfan, H. Jawad, B. B. Felix, S. F. Abbasi, A. Nawaz, S. Akbarzadeh, M. Awais, L. Chen, T. Westerlund, and W. Chen, "Non-wearable IoT-based smart ambient behavior observation system," *IEEE Sensors J.*, vol. 21, no. 18, pp. 20857–20869, Sep. 2021.
- [4] K. R. Darshan and K. R. Anandakumar, "A comprehensive review on usage of Internet of Things (IoT) in healthcare system," in *Proc. Int. Conf. Emerg. Res. Electron., Comput. Sci. Technol. (ICERECT)*, Dec. 2015, pp. 132–136.
- [5] M. Irfan, P. Shun, B. B. Felix, N. Mustafa, S. F. Abbasi, A. Nahli, A. Subasi, T. Westerlund, and W. Chen, "An IoT-based non-contact ECG system: Sole of the feet/hands palm," *IEEE Internet Things J.*, vol. 10, no. 21, pp. 18718–18732, 2023, doi: 10.1109/IJOT.2023.3283037.
- [6] K. A. Allen, "Promoting and protecting infant sleep," *Adv. Neonatal Care*, vol. 12, no. 5, pp. 288–291, 2012.
- [7] M. Willinger, L. S. James, and C. Catz, "Defining the sudden infant death syndrome (Sids): Deliberations of an expert panel convened by the national institute of child health and human development," *Pediatric Pathol.*, vol. 11, no. 5, pp. 677–684, Jan. 1991.

- [8] C. E. Gaw, T. Chounthirath, J. Midgett, K. Quinlan, and G. A. Smith, "Types of objects in the sleep environment associated with infant suffocation and strangulation," *Academic Pediatrics*, vol. 17, no. 8, pp. 893–901, Nov. 2017.
- [9] S. L. Blunden, K. R. Thompson, and D. Dawson, "Behavioural sleep treatments and night time crying in infants: Challenging the status quo," *Sleep Med. Rev.*, vol. 15, no. 5, pp. 327–334, Oct. 2011.
- [10] M. Awais, X. Long, B. Yin, S. F. Abbasi, S. Akbarzadeh, C. Lu, X. Wang, L. Wang, J. Zhang, J. Dudink, and W. Chen, "A hybrid DCNN-SVM model for classifying neonatal sleep and wake states based on facial expressions in video," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 5, pp. 1441–1449, May 2021.
- [11] L. Fraiwan and M. Alkhodari, "Neonatal sleep stage identification using long short-term memory learning system," *Med. Biol. Eng. Comput.*, vol. 58, no. 6, pp. 1383–1391, Jun. 2020.
- [12] E. Alickovic and A. Subasi, "Ensemble SVM method for automatic sleep stage classification," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 6, pp. 1258–1265, Jun. 2018.
- [13] A. R. Hassan and A. Subasi, "A decision support system for automated identification of sleep stages from single-channel EEG signals," *Knowl.-Based Syst.*, vol. 128, pp. 115–124, Jul. 2017.
- [14] J. J. Newson, "The EEG from infancy to adulthood," Sapien Labs, Washington, DC, USA, Tech. Rep., Nov. 2017.
- [15] M. Awais, C. Chen, X. Long, B. Yin, A. Nawaz, S. F. Abbasi, S. Akbarzadeh, L. Tao, C. Lu, L. Wang, R. M. Aarts, and W. Chen, "Novel framework: Face feature selection algorithm for neonatal facial and related attributes recognition," *IEEE Access*, vol. 8, pp. 59100–59113, 2020.
- [16] M. André, M.-D. Lamblin, A. M. d'Allest, L. Curzi-Dascalova, F. Moussalli-Salefranque, S. N. Tich, M.-F. Vecchierini-Blineau, F. Wallois, E. Walls-Esquivel, and P. Plouin, "Electroencephalography in premature and full-term infants. Developmental features and glossary," *Neurophysiologie Clinique/Clin. Neurophysiol.*, vol. 40, no. 2, pp. 59–124, May 2010.
- [17] A. Dereymaeker, K. Pillay, J. Vervisch, M. De Vos, S. Van Huffel, K. Jansen, and G. Naulaers, "Review of sleep-EEG in preterm and term neonates," *Early Hum. Develop.*, vol. 113, pp. 87–103, Oct. 2017.
- [18] S. Montazeri Moghadam, P. Nevalainen, N. J. Stevenson, and S. Vanhatalo, "Sleep state trend (SST), a bedside measure of neonatal sleep state fluctuations based on single EEG channels," *Clin. Neurophysiol.*, vol. 143, pp. 75–83, Nov. 2022.
- [19] A. H. Ansari, O. De Wel, K. Pillay, A. Dereymaeker, K. Jansen, S. Van Huffel, G. Naulaers, and M. De Vos, "A convolutional neural network outperforming state-of-the-art sleep staging algorithms for both preterm and term infants," *J. Neural Eng.*, vol. 17, no. 1, Jan. 2020, Art. no. 016028.
- [20] S. Farooq Abbasi, H. Jamil, and W. Chen, "EEG-based neonatal sleep stage classification using ensemble learning," *Comput., Mater. Continua*, vol. 70, no. 3, pp. 4619–4633, 2022.
- [21] L. Fiorillo, A. Puiatti, M. Papandrea, P.-L. Ratti, P. Favaro, C. Roth, P. Bargiotas, C. L. Bassetti, and F. D. Faraci, "Automated sleep scoring: A review of the latest approaches," *Sleep Med. Rev.*, vol. 48, Dec. 2019, Art. no. 101204.
- [22] M. Awais, H. Ghayvat, and W. Chen, "Face and its features detection during nap," in *Proc. IEEE 23rd Int. Conf. Digit. Signal Process. (DSP)*, Nov. 2018, pp. 1–5.
- [23] M. Awais, X. Long, B. Yin, C. Chen, S. Akbarzadeh, S. F. Abbasi, M. Irfan, C. Lu, X. Wang, L. Wang, and W. Chen, "Can pre-trained convolutional neural networks be directly used as a feature extractor for video-based neonatal sleep and wake classification?" *BMC Res. Notes*, vol. 13, no. 1, pp. 1–6, Dec. 2020.
- [24] A. H. Ansari, K. Pillay, A. Dereymaeker, K. Jansen, S. Van Huffel, G. Naulaers, and M. De Vos, "A deep shared multi-scale inception network enables accurate neonatal quiet sleep detection with limited EEG channels," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 3, pp. 1023–1033, Mar. 2022.
- [25] J. Slavio, *Deep Learning and Artificial Intelligence: A Beginners' Guide to Neural Networks and Deep Learning*. Abhishek Kumar via PublishDrive, 2017.
- [26] A. H. Ansari, O. De Wel, M. Lavanga, A. Caicedo, A. Dereymaeker, K. Jansen, J. Vervisch, M. De Vos, G. Naulaers, and S. Van Huffel, "Quiet sleep detection in preterm infants using deep convolutional neural networks," *J. Neural Eng.*, vol. 15, no. 6, Dec. 2018, Art. no. 066006.
- [27] R. Yu, Z. Zhou, S. Wu, X. Gao, and G. Bin, "MRASleepNet: A multi-resolution attention network for sleep stage classification using single-channel EEG," *J. Neural Eng.*, vol. 19, no. 6, Dec. 2022, Art. no. 066025.
- [28] A. Abbas, S. F. Abbasi, M. Z. Ali, S. Shahid, and W. Chen, "A single channel EEG-based algorithm for neonatal sleep-wake classification," in *Proc. Int. Conf. Adv. Comput. Informat.* Cham, Switzerland: Springer, 2022, pp. 345–352.
- [29] H. Zhu, L. Wang, N. Shen, Y. Wu, S. Feng, Y. Xu, C. Chen, and W. Chen, "MS-HNN: Multi-scale hierarchical neural network with squeeze and excitation block for neonatal sleep staging using a single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 2195–2204, 2023.
- [30] S. F. Abbasi, J. Ahmad, A. Tahir, M. Awais, C. Chen, M. Irfan, H. A. Siddiq, A. B. Waqas, X. Long, B. Yin, S. Akbarzadeh, C. Lu, L. Wang, and W. Chen, "EEG-based neonatal sleep-wake classification using multilayer perceptron neural network," *IEEE Access*, vol. 8, pp. 183025–183034, 2020.
- [31] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.
- [32] H. Phan and K. Mikkelsen, "Automatic sleep staging of EEG signals: Recent development, challenges, and future directions," *Physiological Meas.*, vol. 43, no. 4, Apr. 2022, Art. no. 04TR01.
- [33] M. K. Kiyimik, M. Akin, and A. Subasi, "Automatic recognition of alertness level by using wavelet transform and artificial neural network," *J. Neurosci. Methods*, vol. 139, no. 2, pp. 231–240, Oct. 2004.
- [34] A. Subasi, "Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients," *Exp. Syst. Appl.*, vol. 28, no. 4, pp. 701–711, May 2005.
- [35] A. Subasi, A. Saikia, K. Bagedo, A. Singh, and A. Hazarika, "EEG-based driver fatigue detection using FAWT and multiboosting approaches," *IEEE Trans. Ind. Informat.*, vol. 18, no. 10, pp. 6602–6609, Oct. 2022.
- [36] V. Gupta, M. D. Chopda, and R. B. Pachori, "Cross-subject emotion recognition using flexible analytic wavelet transform from EEG signals," *IEEE Sensors J.*, vol. 19, no. 6, pp. 2266–2274, Mar. 2019.
- [37] J. Kevric and A. Subasi, "Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system," *Biomed. Signal Process. Control*, vol. 31, pp. 398–406, Jan. 2017.
- [38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [39] L. Fraiwan and K. Lweesy, "Neonatal sleep state identification using deep learning autoencoders," in *Proc. IEEE 13th Int. Colloq. Signal Process. Appl. (CSPA)*, Mar. 2017, pp. 228–231.
- [40] V. Gerla, K. Paul, L. Lhotska, and V. Krajca, "Multivariate analysis of full-term neonatal polysomnographic data," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 1, pp. 104–110, Jan. 2009.
- [41] L. Fraiwan and K. Lweesy, "Newborn sleep stage identification using multiscale entropy," in *Proc. 2nd Middle East Conf. Biomed. Eng.*, Feb. 2014, pp. 361–364.



MUHAMMAD IRFAN (Member, IEEE) received the bachelor's degree in electronics and communication engineering from UET Taxila, Pakistan, and the master's degree in electronics and communication engineering from Fudan University, China. He is currently pursuing the dual Ph.D. degree with the Center of Intelligent Medical Electronics, Fudan University, and the Turku Intelligent Embedded and Robotic Systems (TIERS) Group, University of Turku, Finland. He received a grant

from the Chinese Government (CGS) for both his master's and Ph.D. studies. He got a Finnish National Agency for Education Fellowship from Finland to complete his Ph.D. research with the University of Turku. From July 2015 to October 2016, he was with AEDB and Pak Agro Tech, Islamabad, Pakistan, as an Assistant Engineer and a Trainee Engineer. Additionally, he has collaborated with Bear Net and Nile Intelligent companies, Shanghai, on two industrial projects. His research interests include IoMT, edge computing, intelligent sensor systems, sleep analysis, and deep learning.



HAFZA AYESHA SIDDIQA received the B.Sc. degree in electrical engineering from HITEC University Taxila, Pakistan, in 2015, and the master's degree from the Department of Electrical Engineering, HITEC University Taxila. She is currently pursuing the Ph.D. degree in biomedical engineering with Fudan University, China. Her research interests include machine learning, deep learning, fuzzy control, neonatal sleep study, and biomedical signal processing.



ANUM NAWAZ received the M.Sc. (Tech.) degree in information and communication science and technology from the University of Turku, Finland. She is currently pursuing the dual Ph.D. degree with the Shanghai Key Laboratory of Intelligent Information Processing Laboratory, Fudan University, China, and the Turku Intelligent Embedded and Robotic Systems (TIERS) Group, University of Turku. She got a fully-funded scholarship grant from the Shanghai Government Scholarship (SGS) for her Ph.D. degree. Her research interests include information security, the privacy of edge devices, blockchain, ubiquitous healthcare, and autonomous systems.



ABDELWAHED NAHLIIS received the B.S. degree in computer engineering from Hassan II University, Casablanca, in 2015, and the M.S. degree in information and communication engineering from Shanghai University, Shanghai, in 2021. He is currently a Ph.D. Researcher with Fudan University, Shanghai, and a Research Fellow of pattern recognition and data science with the Shanghai Institute. His research interests include the Internet of Things, computer vision, artificial intelligence, and data science.



ABDULHAMIT SUBASI was with many institutions as an Academician. He was with the Georgia Institute of Technology, Georgia, USA, as a Researcher. He was a Professor of computer science with Effat University, Jeddah, Saudi Arabia, from 2015 and 2020. Since 2020, he has been a Professor with the Faculty of Medicine, University of Turku, Turku, Finland. He is specialized in artificial intelligence, machine learning, biomedical signaling, and image analysis. Concerning the application of machine learning to different fields, he wrote more than 30 book chapters and more than 200 published journal and conference papers. He is the author of the books, such as *Practical Guide for Biomedical Signals Analysis Using Machine Learning Techniques*, *Practical Machine Learning for Data Analysis Using Python*, and *Applications of Artificial Intelligence in Medical Imaging*. He received the Queen Effat Award for Excellence in Research, in May 2018.



Her research interests include biomedical engineering, focusing on wearable sensor systems, biomedical signal processing, sleep analysis, and personalized health monitoring.

CHEN CHEN received the M.S. degree in embedded systems from Institut supérieur d'électronique de Paris (ISEP), in 2013, and the Ph.D. degree in computer science from Universit Pierre et Marie Curie (UPMC), in 2016. From 2017 to 2021, she was a Postdoctoral Fellow with the Center for Intelligent Medical Electronics (CIME), School of Information Science and Technology, Fudan University. She is currently an Associate Professor with the Human Phenome Institute, Fudan University.



TOMI WESTERLUND (Senior Member, IEEE) is currently a Professor of robotics and autonomous systems with the University of Turku, Finland. He leads the Turku Intelligent Embedded and Robotic Systems Research Group, University of Turku. His current research interests include the industrial IoT, smart cities and autonomous vehicles (aerial, ground, and surface), and (co-)robots. In all these application areas, the core research interests include energy efficiency, dependability, interoperability, fog/edge computing, and edge AI.



YAN XU received the medical degree from Southern Medical University, in 2008. She is currently a Neurophysiologist with Children's Hospital of Fudan University. She has been a neonatologist for about ten years. She became a Paediatric Neurophysiologist seven years ago, involved mainly on neonatal neurophysiology. To date, she has analyzed thousands of neonatal EEGs.



WEI CHEN (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees from the School of Electronics and Information Engineering, Xian Jiao Tong University, China, in 1999 and 2002, respectively, and the Ph.D. degree from the Department of Electrical and Electronics Engineering, The University of Melbourne, Australia, in 2007. She was with Intern with Bell Laboratories, Germany, and Alcatel-Lucent, Stuttgart, in 2005. She was a Research Assistant with the Department of Electrical and Electronics Engineering, The University of Melbourne, in 2007. From 2007 to 2015, she was an Assistant Professor with the Eindhoven University of Technology, The Netherlands. From 2009 to 2013, she was the Chair of the Theme Health Care, Department of Industrial Design, Eindhoven University of Technology. Since October 2015, she has been a Full Professor and the Director of the Center for Intelligent Medical Electronics (CIME), Department of Electronic Engineering, School of Information Science and Technology, Fudan University. Her research interests include patient health monitoring, medical monitoring system design using wearable sensors, sleep monitoring, brain activity monitoring, wireless body area networks, ambient intelligence, personalized and smart environment, smart sensor systems, and signal processing. She is an Associate Editor of IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING and IEEE JOURNAL ON BIOMEDICAL HEALTH INFORMATICS and a Managing Editor of IEEE REVIEWS IN BIOMEDICAL ENGINEERING.



perinatal brain injury and focusing on wearable sensor systems, sleep analysis, and music therapy.

LAISHUAN WANG received the M.S. degree in paediatrics, in 2000, and the M.D./Ph.D. degree in pediatrics/neonatology from Fudan University, in 2003. He had his perinatal-neonatal fellowship training with UBC, Canada, from 2005 to 2007. He has been with Fudan Children's Hospital, until now. He is currently a Full Professor/Chief Physician of neonatology (Staff Neonatologist) with Children's Hospital of Fudan University. His research interests include the prevention of

...