PPG-Based Sleep Stage Classification Using Pulse Wave Feature Fusion and Explainable AI

Florentin Smarandache
Mathematics, Physics, and Natural Science Division, University of New Mexico, USA
smarand@unm.edu

Satyasri Akula
Department of Data Science, University of Texas at Austin, USA
solutions@sritechtalks.com

Saleh I. Alzahrani
Biomedical Engineering Department, College of Engineering, Imam Abdulrahman Bin Faisal University, Saudi Arabia
sialzahrani@iau.edu.sa

Farrukh Arslan
Riphah School of Computing and Innovation, Riphah International University, Lahore Campus, Pakistan
farrukh.arslan@riphah.edu.pk (corresponding author)

Amir Ijaz
Department of Computing, University of Turku, Finland
amir.ijaz@utu.fi

Received: 30 June 2025 | Revised: 21 July 2025, 9 August 2025, and 17 August 2025 | Accepted: 20 August 2025

 DOI: https://doi.org/10.48084/etasr.13077

ABSTRACT

Sleep monitoring plays a crucial role in understanding and managing various health conditions, including sleep disorders, cardiovascular diseases, and mental health. Traditional sleep monitoring methods rely on Electroencephalography (EEG) and Polysomnography (PSG) in clinical settings. However, these methods are expensive, difficult to administer, and unsuitable for home-based monitoring. In recent years, photoplethysmogram (PPG) has emerged as a promising noninvasive technology that is widely used in wearable devices and holds great potential for sleep assessment. Yet, most current sleep monitoring methods rely on deep learning models, which are inherently "black-box" and challenging in the clinical decision-making process. In this paper, we propose an explainable random forest model for sleep stage classification using pulse wave feature fusion. Our method employs statistical, temporal, and nonlinear dynamical features extracted from the PPG pulse wave associated with sleep patterns. Additionally, we investigate the digital biomarkers of sleep and PPG using SHAP (SHapley Additive exPlanations) methods to enhance interpretability. The proposed approach demonstrates competitive performance, achieving an overall accuracy of 82.56% in two-stage (sleep and wake) classification, 77.79% in three-stage (wake, NREM, REM) classification, and 69.20% in four-stage (wake, light sleep, deep sleep, REM) classification. The results highlight the potential of PPG-based wearable devices in sleep monitoring, offering a feasible solution for home-based assessments with clinical applicability.

Keywords-sleep staging; photoplethysmography; feature fusion; random forest; explainable AI

I. INTRODUCTION

Sleep is marked by inactive sensory organs, minimal muscle movement, and decreased responsiveness to stimuli [1]. The American Academy of Sleep Medicine (AASM) divides sleep into Non-Rapid Eye Movement (NREM) and Rapid Eye Movement (REM) [2]. The NREM is further categorized into N1, N2, and N3 stages. The N1 is the transition from wakefulness to sleep known as light sleep, N2 refers to intermediate sleep, which is characterized by sleep spindles, and N3 is a state of deep sleep known as Slow-Wave Sleep (SWS), crucial for memory consolidation and physical

recovery. In older sleep staging systems, such as the Rechtschaffen and Kales (R&K) scoring system, deep sleep was divided into N3 and N4, both of which were considered SWS due to their high delta wave activity [3]. In 2007, the AASM revised the sleep staging rules, merging the previous N3 and N4 into a single stage called N3 (now simply referred to as "Deep Sleep" or "Slow-Wave Sleep") [2]. REM is the dreaming stage, which is associated with cognitive processing. A typical sleep cycle rounds between NREM and REM stages, repeating throughout the night.

Poor sleep quality has a clinical impact on physical and mental health. Chronic sleep deprivation is linked to a wide range of health issues, including an increased risk of cardiovascular diseases, hypertension, and diabetes. It impairs the body's immune system, making individuals more susceptible to infections. Furthermore, poor sleep is strongly associated with cognitive decline, memory problems, and a decreased ability to concentrate. Sleep disturbances are associated with hormonal balance, leading to weight gain and metabolic issues. Long-term poor sleep quality can accelerate aging and impair decision-making [4]. Thus, monitoring sleep stages is essential for detecting and addressing sleep-related problems. Traditional sleep monitoring methods are based on Electroencephalography (EEG) and Polysomnography (PSG), which are the gold standard in clinical settings for diagnosing and assessing sleep disorders [5]. These methods involve the attachment of multiple electrodes to the scalp and other parts of the body to measure brain activity, muscle movements, eye movements, and other physiological parameters. While highly accurate, these approaches come with significant limitations. One of the primary drawbacks is their relatively high cost. EEG and PSG require specialized equipment, trained personnel to administer the tests, and overnight stays in sleep clinics. The lack of portability and cost of these methods make them impractical for individuals seeking non-disruptive, continuous, and convenient monitoring of their sleep at home. Consequently, there has been a growing demand for more accessible, cost-effective, and noninvasive alternatives that can provide reliable sleep assessments without the need for extensive medical infrastructure.

Photoplethysmography (PPG) is a noninvasive optical technology used to measure changes in blood volume in the microvascular bed of tissue [6]. This technology has become widely available in wearable devices such as smartwatches and fitness trackers, making it both affordable and accessible to the general population. PPG is increasingly being integrated into various healthcare applications due to its ease of use, low cost, and ability to provide continuous monitoring without the need for specialized equipment. There is a growing opportunity to use PPG for sleep monitoring, offering a noninvasive and home-based alternative to traditional sleep monitoring [7]. The clinical association of blood volume with the autonomic nervous system linked sleep with PPG. The advent of Machine Learning (ML) in biomedical applications has revolutionized healthcare systems [8, 9]. A growing research interest has been dedicated to improving PPG-based sleep staging through classical ML techniques. Early studies extracted rhythm-related information, such as inter-beat intervals, pulse-rate variability, and respiratory sinus arrhythmia alongside morphological indices. These features were then fed into conventional classifiers, such as support-vector machine, k-nearest neighbor, random forest, and gradient-boosting ensembles [10-13]. Authors in [10] extracted frequency-domain, time-domain, and nonlinear features from a single-channel PPG signal and used a Light Gradient Boosting Machine to perform multiclass sleep-stage classification, reporting an overall accuracy of 70 % and a Cohen's κ of 0.60. Authors in [11] focused on Heart-Rate-Variability (HRV) analysis, deriving HRV-based features from PPG and applying XGBoost for binary sleep–wake classification, and they achieved an F1-score of 83.9 ± 0.6 % and Cohen's κ of 0.48 ± 0.01. Authors in [12] compiled 72 handcrafted PPG attributes and reached 75.3 % accuracy with a CatBoost ensemble. Authors in [13] combined statistical, morphological, surrogate arterial-pressure, and nonlinear features and obtained 72.4% accuracy using support-vector machines. Notably, none of the above studies incorporated Explainable artificial intelligence (XAI) techniques to explain the model decisions, leaving a gap in trust and transparency for clinical translation. XAI has gained attention for its ability to provide transparency in the decision-making processes of ML methods [14]. To the best of our knowledge, this is the first study to investigate the PPG statistemporal biomarkers associated with sleep patterns using XAI. The key contributions of the current work are:

- We performed statistical, temporal, and non-linear dynamical feature fusion and selected the best features using recursive feature elimination on sleep to blood volume mechanics.
- We developed a grid search cross-validation-based random forest approach for multi-stage sleep classification. We evaluated the proposed approach in 10-fold cross-validation against state-of-the-art and sibling methods.
- Finally, we implemented XAI using SHapley Additive exPlanations (SHAP) to investigate the digital biomarkers associated with sleep patterns, bringing transparency and trust to the decision-making process.

II. METHODOLOGY

Figure 1 presents the framework of the proposed approach for sleep stage classification. The diagram outlines the key modules, including data acquisition, PPG preprocessing, statistemporal feature fusion, feature selection, classification using the Random Forest algorithm, and explainability through XAI. The details of each module are explained in the subsequent sections.

A. Dataset

This study used the acquired data from 10 volunteers (nine male, one female; aged 43–75) from Charité Hospital, Berlin, all diagnosed with sleep-disordered breathing and no cardiac history [13, 15, 16]. The Charité Ethics Committee granted ethical approval, and written informed consent was obtained. Overnight PSG recordings (6.8–10.1 hours) included PPG, EEG, and other biosignals. Following AASM guidelines, the data were segmented into 30-second epochs and manually labeled by an expert, resulting in 9,394 sleep epochs. This work focuses specifically on PPG data for sleep stage classification.

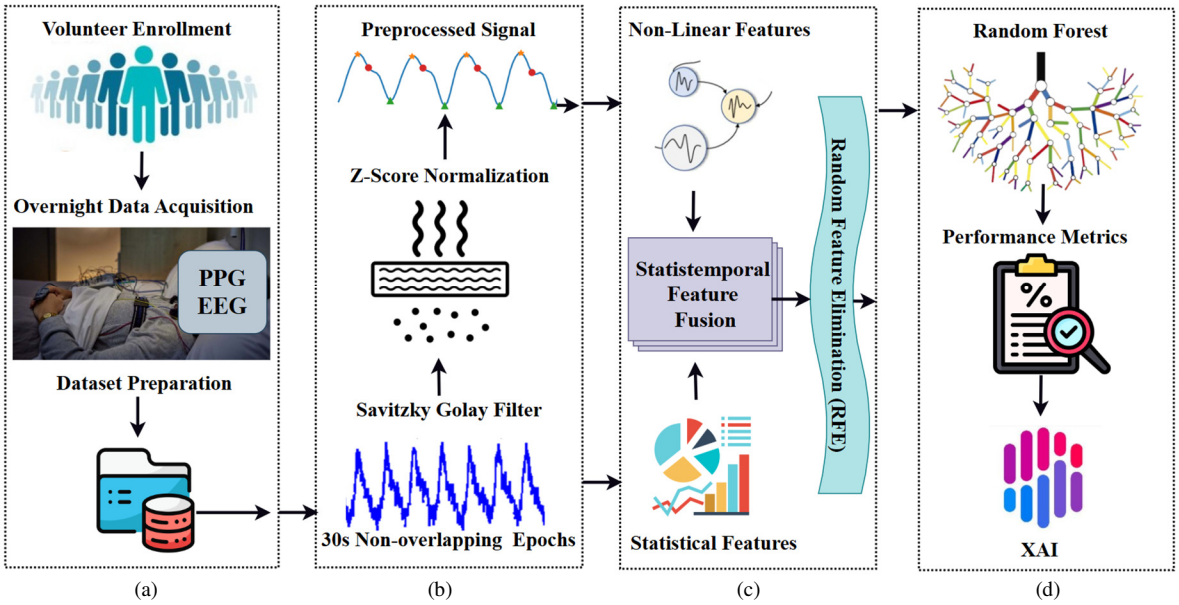

Fig. 1. Framework of the proposed sleep stage classification approach. (a) Data acquisition process which begins from volunteer enrollment to overnight data acquisition and dataset preparation. (b) PPG preprocessing steps from segmentation to filtering, z-score normalization and preprocessed signal creation. (c) The preprocessed signal's nonlinear features and statistical feature extraction with subsequent statistemporal feature fusion process and random feature elimination. (d) Sleep stage classification using RFE based features with Random Forest, and XAI-based interpretation of the decision-making process.

B. PPG Preprocessing

PPG signals are highly susceptible to noise, especially due to subject movements, respiration, and external artifacts. One of the major sources of distortion in PPG signals is baseline wander, which is caused by low-frequency components such as respiratory-induced variations and slow drifts. To eliminate this interference, baseline wander removal was applied, ensuring that the extracted features were not affected by unwanted low-frequency fluctuations. Afterward, a Savitzky-Golay FIR filter was employed to smooth the signal while preserving key waveform characteristics, reducing high-frequency noise, and maintaining the integrity of important signal features. To further enhance consistency across subjects, z-score normalization was applied, standardizing each PPG epoch to have a mean of zero and a variance of one. This normalization step mitigated inter-individual variability and ensured that extracted features were comparable across different subjects. The preprocessed signal was then used for sleep stage classification. For binary sleep-wake classification, sleep stages were categorized into two groups: "Wake" and "Sleep," where all NREM and REM epochs were merged into the sleep category. In the three-stage classification, LS and DS were combined into NREM, while REM and Wake remained as separate stages. For the four-stage classification, all expert-labeled sleep stages were preserved, distinguishing between Wake, LS, DS, and REM.

C. Statiotemporal Feature Fusion

To characterize sleep patterns from PPG signals, we extracted statistical features from the time domain and nonlinear dynamical features, which provide insights into the variability and complexity of the signal. Statistical features capture essential descriptive properties of the PPG waveform, while nonlinear features analyze the intrinsic dynamics and irregularity of the signal, making them particularly relevant for sleep stage classification.

- Mean Absolute Deviation (MAD): Measures the average deviation of PPG values from the mean, reflecting the overall variability.
- Median Absolute Deviation (MABD): A robust measure of variability, showing how PPG values deviate from the median, less sensitive to outliers than the standard deviation.
- Interquartile Range (IQR): The difference between the 75th and 25th percentiles of PPG values, representing signal dispersion and variability.
- N^{th} Central Moment (NCM): Captures higher-order moments of the signal distribution, providing insights into the shape and distribution of PPG fluctuations.
- Average Curve Length (ACL): Computes the cumulative absolute differences of PPG values over time, measuring waveform complexity.
- Shape Factor (SF): The ratio of the RMS value to the mean, describing the overall shape of the PPG waveform.
- Mean Value (ME): Represents the average amplitude of the PPG signal.
- Standard Deviation (STD): Quantifies the dispersion of PPG values around the mean, indicating overall signal variability.
- Root Mean Square (RMS): Measures the square root of the mean squared values, capturing signal energy.

- Trimmed Mean (TME25, TME50): Mean values computed after excluding the lower and upper 25% or 50% of PPG values, reducing the effect of extreme values.
- Geometric Mean (GME): Provides an alternative measure of central tendency, useful for skewed distributions.
- Maximum Value (MAX): The highest recorded amplitude in a PPG epoch, indicative of peak signal fluctuations.
- Minimum Value (MIN): The lowest recorded amplitude in a PPG epoch.
- Skewness (SK): Measures the asymmetry of the PPG distribution, indicating whether the signal is biased toward higher or lower values.
- Kurtosis (KU): Describes the sharpness or flatness of the PPG distribution compared to a normal distribution.
- Poincare SD1 (PSD1): Measures short-term variability of the PPG signal in a Poincare plot, capturing fast fluctuations in blood volume changes.
- Poincare SD2 (PSD2): Represents long-term variability in the PPG signal, providing insights into sustained fluctuations.
- Ratio of SD1/SD2 (SD1RSD2): Quantifies the balance between short-term and long-term variability, helping to distinguish different sleep stages.
- Complex Correlation Measure (CCM): Evaluates the degree of complexity and correlation within the PPG signal, reflecting autonomic nervous system activity.
- Hjorth Mobility (HjM): Measures the rate of change in the PPG signal, indicating its smoothness and frequency characteristics.
- Hjorth Complexity (HjC): Assesses the irregularity of the PPG waveform, capturing how signal patterns evolve.
- Higuchi Fractal Dimension (HFD): Quantifies the fractal complexity of the PPG signal, providing a measure of self-similarity and irregularity.
- Katz Fractal Dimension (KFD): Another fractal-based measure that estimates the structural complexity of the PPG waveform, which varies across different sleep stages.

To enhance the discriminative power of sleep stage classification, we employed statiotemporal feature fusion by integrating statistical time-domain features with nonlinear dynamical features, capturing both signal variability and complexity. Feature selection is crucial in optimizing model performance by reducing dimensionality, improving computational efficiency, and eliminating redundant or irrelevant features. This study employed the Recursive Feature Elimination (RFE) method to select the 10 most correlated features from the extracted statistemporal feature set. RFE works by iteratively training a machine learning model and ranking features based on their importance. Initially, the model is trained using all available features, and the least important feature is removed based on its contribution to the classification performance. The final set of selected features retains the most relevant information while reducing redundancy as expressed in (1):

$$STFF = \{MABD; NCM; SF; ME; TME25; GME; MAX; MIN; HjC, HFD \quad (1)$$

We employed Random Forest (RF) as our base model after extensive ablation studies with sibling classifiers for sleep stage classification. RF constructs an ensemble of decision trees using bootstrap aggregation (bagging) and randomly selects features at each split to enhance generalization and reduce overfitting. Hyperparameters were optimized via stratified 5-fold grid search cross-validation over n_estimators {50, 100, 200}, max_depth {10, 20, 30, None}, min_samples_split {2, 5, 10}, and min_samples_leaf {1, 2, 4}. The identified optimal parameters were: n_estimators=200, max_depth=30, min_samples_split=2, min_samples_leaf=1. To provide transparency and interpretability in clinical settings, we employed SHapley Additive exPlanations (SHAP), an XAI method that calculates feature importance through Shapley values. SHAP identified critical statistical and nonlinear PPG features influencing sleep stage classification, thereby offering interpretable insights into digital sleep biomarkers.

III. RESULTS AND DISCUSSION

We conducted extensive ablation studies to evaluate the performance of the proposed Grid Search Cross-Validation Random Forest (GSCV-RF) method against sibling machine learning methods, including Support Vector Machine (SVM), k-nearest Neighbor (KNN), and Logistic Regression (LR), for two-stage sleep classification using 10-fold cross-validation.

The GSCV-RF method outperformed all other classifiers, achieving an overall accuracy of 82.56% and an F1-score of 75.40%. Comparatively, the SVM achieved an accuracy of 71.79% and an F1-score of 69.09%, while KNN reported an accuracy of 76.03% with an F1-score of 75.69%. LR recorded the lowest performance among the tested methods, achieving an accuracy of 64.05% and an F1-score of 64.71%. Further, we employed GSCV-RF for three-stage and four-stage sleep classification using 10-fold cross-validation. For three-stage sleep classification, the proposed method demonstrated competitive results with an accuracy of 77.79%, sensitivity of 62.63%, specificity of 84.77%, and an F1-score of 75.40%. In the more challenging four-stage sleep classification, the proposed approach maintained robust performance, achieving an accuracy of 69.20%, sensitivity of 59.55%, specificity of 87.88%, and an F1-score of 68.10%. Detailed results for each stage using 10-fold cross-validation are summarized in Table I. Figure 2 presents the confusion matrices for the two, three, and four-stage experiments.

The SHAP summary plot in Figure 3 indicates that five features dominate the model's decisions, and each aligns with a well-established physiological hallmark of a specific sleep state. The Higuchi Fractal Dimension *(higFracDim)* captures beat-to-beat temporal irregularity; high values are typical of wakefulness and light sleep because sympathetic activation, muscle tone, and motion artefacts increase PPG complexity. HjC, a measure of rapid spectral fluctuation, peaks in stage

N1/N2, where cortical and autonomic activity are partially desynchronized. The *geometricMean, centralMoment,* and *shapeFactor* rise during slow-wave (deep) sleep, mirroring the large, low-frequency oscillations produced by pronounced vasodilation and reduced heart rate variability. In contrast, moderate positive SHAP contributions of *tmean25* and *medianAbsDev* to REM reflect the burst-like blood-flow changes driven by phasic sympathetic surges. Extreme amplitude excursions captured by *maxValue* and *minValue* distribute across all classes, suggesting that sporadic artefacts or posture shifts can influence every stage, albeit with lower overall importance. Taken together, complexity indices separate wakefulness from sleep, amplitude-shape metrics distinguish deep from light stages, and finer variability indices isolate REM, thereby providing a physiologically coherent explanation for the classifier's four-stage performance.

Table II presents a comparison of the proposed approach with state-of-the-art methods. The proposed method achieved superior results in terms of accuracy while also providing clinical explainability. A LightGBM model achieved 70% accuracy on 27,333 sleep epochs from 27 subjects [10], whereas the logistic regression model attained 74% accuracy on data from 1,831 subjects [11]. On the same dataset of 9,394 sleep epochs from 10 subjects, our method achieved 82.56% accuracy, outperforming CatBoost [12] with 75.29% accuracy and SVM [13] with 72.36% accuracy.

TABLE I. TWO-, THREE-, AND FOUR-CLASS SLEEP STAGE CLASSIFICATION PERFORMANCE EVALUATION IN TERMS OF ACCURACY (ACC), SENSITIVITY (SEN), SPECIFICITY (SPE), AND F1-SCORE (F1-S) WITH 10-FOLD CROSS-VALIDATION

Fold	Two-Stage Sleep Classification				Three-Stage Sleep Classification				Four-Stage Sleep Classification			
	ACC (%)	SEN (%)	SPE (%)	F1-S (%)	ACC (%)	SEN (%)	SPE (%)	F1-S (%)	ACC (%)	SEN (%)	SPE (%)	F1-S (%)
1	83.24	80.77	80.77	76.36	77.62	62.58	85.55	76.36	69.78	61.68	88.22	69.16
2	82.82	80.51	80.51	76.36	77.84	62.07	85.71	76.36	71.16	60.45	88.73	69.90
3	80.06	76.82	76.82	74.72	75.93	61.81	84.15	74.72	68.29	59.33	87.74	67.54
4	82.82	80.12	80.12	76.59	77.52	64.42	85.55	76.59	69.57	60.34	88.17	68.77
5	80.70	77.64	77.64	76.13	76.99	64.51	85.09	76.13	68.82	59.79	87.78	68.05
6	82.93	80.46	80.46	75.91	76.99	63.19	85.11	75.91	70.39	60.33	88.43	69.37
7	81.44	78.60	78.60	75.58	76.78	62.62	84.74	75.58	68.50	59.6	87.59	67.77
8	82.71	79.53	79.53	72.76	74.02	59.55	83.19	72.76	66.28	56.64	86.88	65.34
9	82.91	80.42	80.42	76.15	77.07	64.00	85.23	76.15	69.53	58.35	87.97	68.26
10	80.04	76.83	76.83	73.45	74.42	61.58	83.35	73.45	67.52	58.97	87.30	66.87
Overall	**82.56**	**79.17**	**79.17**	**75.40**	**77.79**	**62.63**	**84.77**	**75.40**	**69.20**	**59.55**	**87.88**	**68.10**

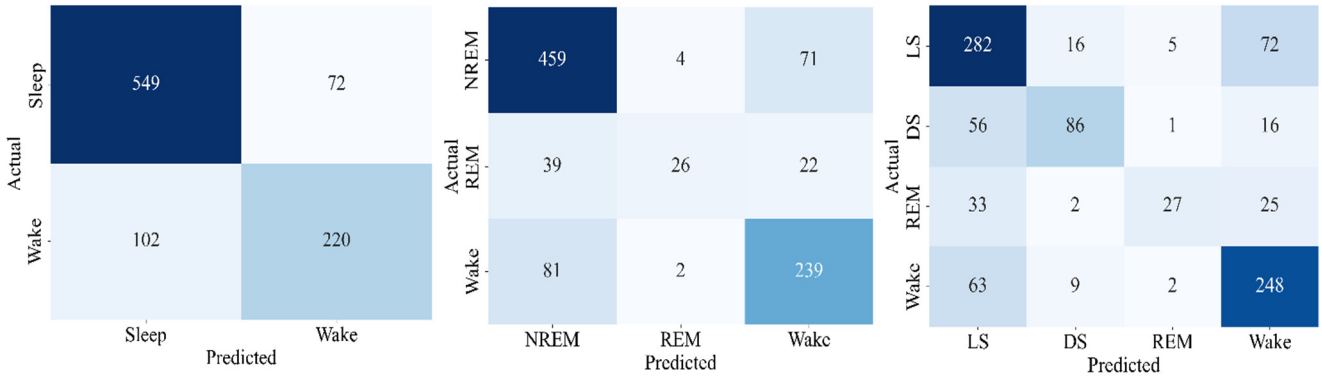

Fig. 2. Confusion matrix for two-, three-, and four-stage sleep classification (left to right)

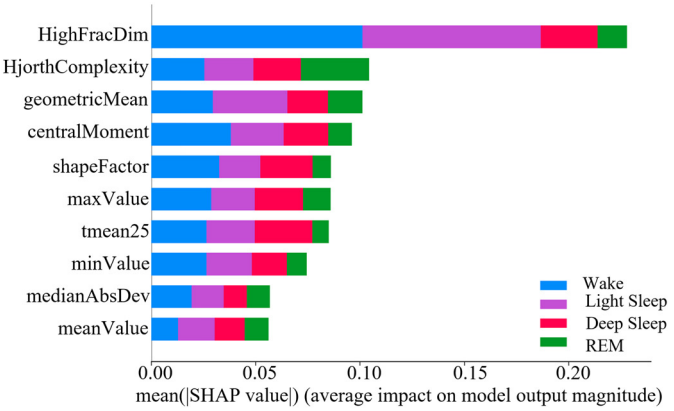

Fig. 3. Summary plot of SHAP analysis for feature importance.

TABLE II. COMPARISON WITH PREVIOUS WORKS

Reference	Method	Sleep Epochs	Explainability	Accuracy (%)
[10]	LightGBM	27,333	Absent	70.00
[11]	LR	1831	Absent	74.00
[12]	CatBoost	9,394	Absent	75.29
[13]	SVM	9,394	Absent	72.36
Our	Random Forest	9,394	Applied	82.56

Most of the prior PPG sleep-staging studies did not explain model decisions, limiting their clinical utility. To thee best of our knowledge, this is the first XAI-based analysis of statistical temporal PPG biomarkers of sleep. SHAP reveals stage-specific drivers: Higuchi fractal dimension and Hjorth complexity for wake/light sleep; geometric mean, central moment, and shape factor for slow-wave sleep; and tmean25 and median absolute deviation for REM. These

explanations link predictions to blood-volume mechanics and provide trustworthy evidence for clinical deployment.

IV. CONCLUSION

In this work, an interpretable sleep stage classification method on wearable PPG signals was developed and evaluated. We employed the statistical, temporal, and nonlinear dynamical features associated with sleep patterns. Subsequently, a random forest model with grid search cross-validation was developed for 2-stage, 3-stage, and 4-stage sleep classification. The proposed approach achieved accuracies of 82.56%, 77.79%, and 69.20%; sensitivities of 79.17%, 62.63%, and 59.55%; specificities of 79.17%, 84.77%, and 87.88%; and F1-scores of 75.40%, 75.40%, and 68.10% for two-stage, three-stage, and four-stage sleep classification, respectively. Finally, we performed a Shapley additive explanation to interpret the decision-making process of the proposed approach, which is very important in clinical systems. Our analysis revealed that the Higuchi Fractal Dimension, Hjorth Complexity, geometric Mean, central Moment, and Shape Factor are the most influential features driving accurate four-stage sleep classification. These descriptors collectively reflect waveform complexity, spectral variability, and amplitude symmetry, and thus provide physiologically meaningful insight into each sleep state. In future work, we plan to broaden the study population by incorporating multi-regional datasets and to investigate additional clinically relevant PPG biomarkers in larger cohorts.

REFERENCES

[1] L. N. Whitehurst, A. Subramoniam, A. Krystal, and A. A. Prather, "Links between the brain and body during sleep: implications for memory processing," *Trends in Neurosciences*, vol. 45, no. 3, pp. 212–223, Mar. 2022, https://doi.org/10.1016/j.tins.2021.12.007.

[2] R. B. Berry *et al.*, "AASM Scoring Manual Updates for 2017 (Version 2.4)," *Journal of Clinical Sleep Medicine*, vol. 13, no. 05, pp. 665–666, https://doi.org/10.5664/jcsm.6576.

[3] D. Moser *et al.*, "Sleep classification according to AASM and Rechtschaffen & Kales: effects on sleep scoring parameters," *Sleep*, vol. 32, no. 2, pp. 139–149, Feb. 2009, https://doi.org/10.1093/sleep/32.2.139.

[4] V. K. Chattu, Md. D. Manzar, S. Kumary, D. Burman, D. W. Spence, and S. R. Pandi-Perumal, "The Global Problem of Insufficient Sleep and Its Serious Public Health Implications," *Healthcare*, vol. 7, no. 1, Dec. 2018, https://doi.org/10.3390/healthcare7010001.

[5] X. Zhang, X. Zhang, Q. Huang, Y. Lv, and F. Chen, "A review of automated sleep stage based on EEG signals," *Biocybernetics and Biomedical Engineering*, vol. 44, no. 3, pp. 651–673, Jul. 2024, https://doi.org/10.1016/j.bbe.2024.06.004.

[6] J. Allen, D. Zheng, P. A. Kyriacou, and M. Elgendi, "Photoplethysmography (PPG): state-of-the-art methods and applications," *Physiological Measurement*, vol. 42, no. 10, Aug. 2021, Art. no. 100301, https://doi.org/10.1088/1361-6579/ac2d82.

[7] T.-H. Tran, P. A. Nguyen, L. A. Ngoc, D.-T. Tran, and M. T. Pham, "Optimal CNN Model for Obstructive Sleep Apnea Detection using Particle Swarm Optimization," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19553–19560, Feb. 2025, https://doi.org/10.48084/etasr.9154.

[8] A. T, H. Grace, N. Martin, and F. Smarandache, "Enhanced Neutrosophic Set and Machine Learning Approach for Breast Cancer Prediction," *Neutrosophic Sets and Systems*, vol. 73, no. 1, Sep. 2024.

[9] A. Sufian, A. Ghosh, A. S. Sadiq, and F. Smarandache, "A Survey on Deep Transfer Learning to Edge Computing for Mitigating the COVID-19 Pandemic," *Journal of Systems Architecture*, vol. 108, Sep. 2020, Art. no. 101830, https://doi.org/10.1016/j.sysarc.2020.101830.

[10] X. Zhao and G. Sun, "A Multi-Class Automatic Sleep Staging Method Based on Photoplethysmography Signals," *Entropy*, vol. 23, no. 1, Jan. 2021, Art. no. 116, https://doi.org/10.3390/e23010116.

[11] D. A. Almeida *et al.*, "A machine-learning sleep-wake classification model using a reduced number of features derived from photoplethysmography and activity signals," in *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, Jun. 2024, pp. 61–69, https://doi.org/10.5753/sbcas.2024.1872.

[12] T. Ferdous, R. U. Karim, A. Samin, S. Mahdi, H. Tasnim, and A. N. Zereen, "Machine Learning Approaches in Photoplethysmography-Based Sleep Stage Classification," in *2024 IEEE 2nd International Conference on Electrical, Automation and Computer Engineering (ICEACE)*, Sep. 2024, pp. 123–128, https://doi.org/10.1109/ICEACE63551.2024.10898858.

[13] M. A. Motin, C. Kumar Karmakar, T. Penzel, and M. Palaniswami, "Sleep-Wake Classification using Statistical Features Extracted from Photoplethysmographic Signals," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, Jul. 2019, pp. 5564–5567, https://doi.org/10.1109/EMBC.2019.8857761.

[14] J. Shen *et al.*, "Physiological signal analysis using explainable artificial intelligence: A systematic review," *Neurocomputing*, vol. 618, Feb. 2025, Art. no. 128920, https://doi.org/10.1016/j.neucom.2024.128920.

[15] M. A. Motin, "Matlab Code," *figshare*, Jan. 31, 2023. https://figshare.com/articles/code/Matlab_Code/21981983/1.

[16] M. Abdul Motin, C. Kamakar, P. Marimuthu, and T. Penzel, "Photoplethysmographic-based automated sleep-wake classification using a support vector machine," *Physiological Measurement*, vol. 41, no. 7, Aug. 2020, Art. no. 075013, https://doi.org/10.1088/1361-6579/ab9482.