Interpretable Feature Based Machine Learning for Automatic Sleep Detection Using Photoplethysmography: A Cross Disciplinary Approach to Health Management

1st Kaiwen Xiao
School of arts and design
Yanshan university
Qinhuangdao, China
13718031392@163.com

2nd Zhaoliang Huang
Guangzhou Songyin Electronic Technology Co., Ltd.
Guangzhou, China
qq903251603@163.com

3rd Ruifeng Xie *
School of Art and Design
GuangZhou City Construction College
Guangzhou, China
13922177828@139.com

***Abstract*—Sleep disorders pose a significant global health challenge, impacting millions and leading to a myriad of adverse health outcomes. Accurate and accessible sleep wake classification is paramount for effective diagnosis, management, and the promotion of overall well being. While polysomnography (PSG) remains the gold standard, its high cost, invasiveness, and logistical complexities limit its widespread applicability. Photoplethysmography (PPG), a non invasive and cost effective optical technique, has emerged as a promising alternative for continuous sleep monitoring. However, existing PPG based sleep detection methods often lack interpretability, struggle with class imbalance, and rarely integrate cross disciplinary insights from design, engineering, business, and culture to enhance user experience and practical utility. This paper presents a novel interpretable feature based machine learning framework for automatic sleep detection using PPG, specifically designed to bridge these gaps. Leveraging a comprehensive dataset and employing advanced signal processing and feature engineering techniques, our model achieves robust sleep wake classification. Crucially, we integrate principles from user centered design to ensure intuitive data visualization, explore innovative business models for scalable deployment, and consider cultural nuances to foster broader adoption. Our findings demonstrate that a cross disciplinary approach not only improves the technical performance of PPGbased sleep detection but also significantly enhances its clinical applicability and societal impact, paving the way for more personalized and effective sleep health management solutions.**

Keywords—Photoplethysmography (PPG), Sleep–wake classification, Interpretable machine learning, Feature engineering, Class imbalance (ADASYN), Random Forest, Wearable sensors

1. INTRODUCTION

Sleep, a fundamental physiological process, is indispensable for maintaining optimal physical and mental health, cognitive function, and overall quality of life [1]. Despite its critical importance, a substantial portion of the global population suffers from various sleep deficiencies and disorders. For instance, in the United States, approximately one third of adults report insufficient sleep [2], while globally, conditions such as obstructive sleep apnea affect an estimated 936 million individuals [3]. These pervasive sleep issues are strongly correlated with a range of severe health problems, including but not limited to obesity, type 2 diabetes, cardiovascular diseases, and mental health disorders suchates depression and anxiety [4]. Consequently, accurate and timely sleep wake detection is not merely a diagnostic tool but a cornerstone for understanding, managing, and mitigating the profound health consequences associated with disturbed sleep patterns.

Traditionally, polysomnography (PSG) has been the undisputed clinical gold standard for comprehensive sleep monitoring. PSG involves the simultaneous recording of multiple physiological parameters, including electroencephalography (EEG), electrooculogram (EOG), electromyogram (EMG), breathing effort, airflow, pulse oximetry, and blood oxygen saturation [5]. The American Academy of Sleep Medicine (AASM) provides a widely accepted framework for sleep staging based on PSG data, categorizing sleep into distinct stages: N1, N2, N3 (deep sleep), and rapid eyemovement (REM) sleep [6]. While PSG offers unparalleled precision and a holistic view of sleep architecture, its inherent limitations—including high cost, time consuming procedures, the requirement for an overnight stay in a specialized laboratory, professional supervision, and manual data labeling—severely restrict its practicality and scalability for routine monitoring and large scale population studies [7]. Furthermore, PSG is susceptible to inter rater variability in manual scoring, which can impact diagnostic consistency [8].

In response to the limitations of PSG, there has been a growing inter est in developing non invasive, cost effective, and user friendly alternatives for sleep monitoring. Photoplethy smography (PPG) stands out as a particularly promising technology in this regard. PPG is an optical technique that measures volumetric changes in blood circulation in the peripheral vasculature typically by illuminating the skin and detecting variations in light absorption or reflection [9]. The resulting PPG waveform reflects changes in blood volume with each heartbeat, providing valuable insights into cardiovascular dynamics and autonomic nervous system activity, which are known to vary across different sleep stages [10][11]. The widespread availability of PPG sensors in consumer wearables (e.g., smartwatches, fitness trackers) and its established clinical use for cardiovascular monitoring make it an attractive candidate for scalable and continuous home based sleep assessment.

Despite the significant advancements in PPG based sleep detection, several critical challenges persist. Many existing models, particularly those based on deep learning, often operate as black boxes, lacking the interpretability crucial for clinical adoption and trust [12]. Furthermore, the inherent class imbalance in sleep data (e.g., wake periods constituting a smaller proportion than sleep stages) can bias models towards majority classes, leading to suboptimal performance in detecting less prevalent but clinically significant states [13]. Most importantly, current research often focuses solely on technical performance metrics, overlooking the broader implications of integrating such technology into real world health management ecosystems. There is a notable gap in research that systematically incorporates cross disciplinary perspectives—from design, engineering, business, and culture—to create truly impactful and user centric sleep health solutions.

This paper aims to address these critical gaps by developing an interpretable featurebased machine learning framework for automatic sleep detection using PPG, specifically from a novel cross disciplinary innovation perspective. Our primary objectives are:

- To develop a robust and interpretable feature based machine learning model for accurate sleep wake classification using PPG data. This involves advanced signal processing, comprehensive feature engineering, and the application of machine learning algorithms tailored to handle class imbalance and enhance model transparency.
- To integrate principles from user centered design and service design to transform raw physiological data into intuitive and actionable sleep health insights. This includes designing user friendly interfaces for data visualization and personalized intervention strategies that promote user engagement and adherence.

To explore innovative business models and technological frameworks for the scalable deployment and commercialization of PPG based sleep health solutions. This encompasses considering the economic viability, market opportunities, and data security aspects within the digital health landscape. To investigate the cultural adaptability of PPG based sleep management systems, ensuring their relevance and effectiveness across diverse cultural contexts. This involves understanding cultural perceptions of sleep and health, and tailoring solutions to align with varying lifestyles and values.By systematically integrating insights from design, engineering, business, and culture, this research seeks to move beyond purely technical advancements. We envision a future where PPG based sleep detection is not only technically superior but also seamlessly integrated into daily life, empowering individuals to proactively manage their sleep health. This cross disciplinary approach not only contributes significantly to the field of biosensing and machine learning but also offers profound implications for public health, digital health innovation, and the broader intersection of technology and human well being.

2. Related Work

The landscape of sleep monitoring has evolved significantly, driven by advancements in sensor technology, signal processing, and machine learning. This section provides a comprehensive review of existing research pertinent to PPG based sleep detection, critically examining the state of the art in technical methodologies and exploring the nascent integration of cross disciplinary perspectives.

2.1. PPG Based Sleep Detection Techniques

Photoplethysmography (PPG) has gained considerable traction as a noninvasive and continuous method for sleep monitoring, primarily due to its ability to capture cardiovascular dynamics that correlate with sleep stages. Early approaches to PPG based sleep detection largely focused on extracting handcrafted features from the PPG waveform and its derivatives, followed by classification using traditional machine learning algorithms. These features typically include statistical measures (e.g., mean, standard deviation, skewness, kurtosis), frequency domain components (e.g., power spectral density in different bands, reflecting autonomic nervous system activity), and morphological parameters (e.g., pulse transit time, pulse wave velocity, amplitude, width) [14][15].

For instance, studies have utilized features derived from heart rate variability (HRV) and pulse rate variability (PRV), both closely related to PPG, to differentiate between wakefulness and various sleep stages. These features often include time domain metrics (e.g., SDNN, RMSSD), frequency domain metrics (e.g., LF/HF ratio, VLF power), and non linear dynamics (e.g., Poincaré plot indices, sample entropy) [16][17]. Classification algorithms such as Support Vector Machines (SVMs), Random Forests (RFs), K Nearest Neighbors (KNNs), and decision trees have been widely employed, demonstrating varying degrees of success in binary sleep wake classification and multi class sleep staging [18][19]. The primary advantage of feature based approaches lies in their interpretability, allowing researchers and clinicians to understand which physiological parameters contribute most to the classification decision. This transparency is crucial for clinical validation and trust.

More recently, the advent of deep learning has revolutionized PPG based sleep detection, enabling end to end learning directly from raw PPG signals or minimally processed data. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), particularly Long Short Term Memory (LSTM) networks, have shown promise in automatically extracting complex, hierarchical features from time series PPG data, often outperforming traditional feature based methods in terms of raw accuracy [20][21]. For example, some studies have employed CNNs to capture local patterns in PPG segments, while LSTMs are used to model temporal dependencies across longer durations, thereby improving the accuracy of sleep stage transitions. Hybrid models combining CNNs and LSTMs have also been proposed to leverage both spatial and temporal features effectively [22][23]. While deep learning models offer superior performance in many cases, their black box nature often limits their interpretability, making it challenging to understand the underlying physiological mechanisms driving their predictions. This lack of transparency can be a significant barrier to clinical adoption, where understanding the 'why' behind a diagnosis is as important as the 'what'.

Another persistent challenge in PPG based sleep detection is the inherent class imbalance, particularly between wake and sleep stages, and among different sleep stages themselves. Wake periods typically constitute a smaller proportion of a 24 hour cycle compared to sleep, leading to models biased towards the majority class. Techniques such as oversampling (e.g., Synthetic Minority

Over sampling Technique SMOTE, Adaptive Synthetic Sampling ADASYN), under sampling, and cost sensitive learning have been employed to mitigate this issue, with varying degrees of success [24][25]. However, the effectiveness of these balancing techniques often depends on the specific dataset and model architecture.

2.2. Cross Disciplinary Integration in Health Management

While technical advancements in PPG based sleep detection are crucial, the successful translation of these technologies into effective real world health management solutions necessitates a broader, cross disciplinary perspective. Traditionally, health technology research has been dominated by engineering and medical sciences. However, there is a growing recognition that integrating insights from design, business, and cultural studies can significantly enhance the utility, adoption, and impact of health interventions.

Design Perspective: User centered design (UCD) and service design principles are increasingly being applied in digital health to create intuitive, engaging, and effective health technologies. This involves understanding user needs, behaviors, and contexts through ethnographic research, usability testing, and iterative prototyping [26]. For sleep health, design thinking can transform complex physiological data into easily understandable visualizations, personalized feedback, and actionable recommendations. For example, designing a mobile application that presents sleep data in a clear, graphical format, or developing nudges based on behavioral science to encourage healthier sleep habits, can significantly improve user engagement and adherence [27]. The aesthetic and functional design of wearable devices themselves also plays a crucial role in user acceptance and continuous usage.

Business Perspective: The commercial viability and scalability of health technologies depend heavily on robust business models. This involves identifying target markets, value propositions, revenue streams, and key partnerships [28]. For PPG based sleep solutions, business innovation can range from direct to consumer wearable devices with subscription based analytics to B2B models providing sleep monitoring services to corporate wellness programs, insurance companies, or healthcare providers. Understanding regulatory landscapes, data privacy concerns (e.g., GDPR, HIPAA), and reimbursement policies are also critical for successful market entry and sustained growth [29]. The shift towards value based care and preventive health further emphasizes the need for business models that incentivize continuous health monitoring and proactive interventions.

Cultural Perspective: Health behaviors and perceptions are deeply embedded in cultural contexts. What constitutes good sleep, how sleep disorders are perceived, and the willingness to adopt new health technologies can vary significantly across different cultures [30]. For instance, traditional beliefs about health and illness, family structures, daily routines, and dietary habits can all influence sleep patterns and the acceptance of digital health interventions. A culturally sensitive approach to PPGbased sleep detection involves tailoring communication strategies, intervention recommendations, and even product features to resonate with specific cultural values and norms. Ignoring cultural nuances can lead to low adoption rates and ineffective interventions, even if the underlying technology is technically sound [31]. For example, in some cultures, sharing personal health data might be viewed differently than in others, necessitating careful consideration of privacy and data governance frameworks that are culturally appropriate.

2.3. Limitations of Existing Research and the Present Study's Contribution

While the existing body of research has significantly advanced PPG based sleep detection, several critical limitations persist. Firstly, many studies, particularly those employing deep learning, suffer from a lack of interpretability, making it difficult to ascertain the physiological basis of their predictions and hindering clinical trust and adoption. Secondly, the pervasive issue of class imbalance in sleep datasets is often inadequately addressed, leading to models that perform poorly on minority classes, such as wakefulness or specific sleep disorder stages. Thirdly, and most importantly, the vast majority of research remains siloed within technical domains, neglecting the broader implications of integrating such technology into real world health management ecosystems. There is a notable gap in research that systematically incorporates crossdisciplinary perspectives—from design, engineering, business, and culture—to create truly impactful and user centric sleep health solutions.

3. METHODOLOGY AND SYSTEM DESIGN

Our proposed framework for interpretable feature based machine learning for automatic sleep detection using PPG integrates advanced signal processing, comprehensive feature engineering, and robust classification techniques, all within a cross disciplinary innovation paradigm. The overall experimental flowchart is depicted in Figure 1, illustrating the sequential steps from raw PPG signal acquisition to interpretable sleep wake classification.

3.1. Data Acquisition and Preprocessing

Raw PPG signals are typically acquired from wearable devices, such as smartwatches or dedicated PPG sensors. These signals are often contaminated by various forms of noise, including motion artifacts, baseline wander, and powerline interference. Therefore, a multi stage preprocessing pipeline is essential to ensure signal quality and reliability.

- Noise Reduction: A band pass filter (e.g., 0.5 10 Hz) is applied to remove highfrequency noise and low frequency baseline wander, which can be caused by respiration or body movements.
- Artifact Removal: Advanced algorithms, such as adaptive filtering or wavelet based denoising, are employed to mitigate motion artifacts, which are a common challenge in wearable sensor data. Segments heavily corrupted by artifacts are identified and either removed or interpolated.
- Signal Segmentation: The cleaned PPG signal is segmented into fixedlength epochs (e.g., 30 second intervals), consistent with standard sleep staging protocols, to facilitate feature extraction and subsequent classification.

3.2. Feature Extraction

From each preprocessed PPG segment, a comprehensive set of features is extracted, encompassing time domain, frequency domain, and non linear dynamics. These features are carefully selected to capture various physiological aspects related to cardiovascular activity and autonomic nervous system regulation, which are known to change across sleep wake states. A total of 330 features are extracted, including:

- Time Domain Features: Statistical measures of the PPG waveform (e.g., mean, standard deviation, skewness, kurtosis, median, range), pulse interval (PPI) statistics (e.g., mean PPI, standard deviation of PPI, root mean square of successive differences RMSSD), and morphological features (e.g., pulse width at different amplitudes, peak to peak interval, systolic upstroke time, diastolic decay time).
- Frequency Domain Features: Power spectral density (PSD) analysis of the PPG and PPI signals in various frequency bands (e.g., Very Low Frequency (VLF), Low Frequency (LF), High Frequency (HF). Ratios such as LF/HF power are particularly important as they reflect sympathovagal balance [16].
- Non Linear Dynamics Features: Measures that quantify the complexity and predictability of the PPG and PPI signals, such as Sample Entropy, Approximate Entropy, Detrended Fluctuation Analysis (DFA), and Poincaré plot indices (e.g., SD1, SD2, SD1/SD2 ratio). These features can capture subtle changes in physiological control mechanisms that are not apparent from linear analysis.

3.3. Feature Selection

Given the large number of extracted features (330 features), feature selection is crucial to reduce dimensionality, mitigate overfitting, and improve model interpretability. We employ a combination of filter and wrapper methods for optimal feature subset selection.

- Filter Methods: Initial feature ranking is performed using statistical measures such as mutual information, Pearson correlation, and ANOVA Fvalue to identify features with high relevance to the sleep wake labels.
- Wrapper Methods: Recursive Feature Elimination (RFE) with a Random Forest classifier is used to iteratively select the most informative features. This process is performed separately for the unbalanced and ADASYNbalanced datasets to identify distinct feature sets that are most predictive in each scenario.

This rigorous feature selection process results in a refined set of 75 features for the unbalanced dataset and 35 features for the ADASYN balanced dataset, ensuring that only the most relevant and non redundant features are used for classification.

3.4. Class Imbalance Handling: ADASYN Sampling

Class imbalance is a pervasive issue in sleep wake classification, where wake epochs are typically less frequent than sleep epochs. To address this, we employ the Adaptive Synthetic (ADASYN) sampling approach. ADASYN is an advanced oversampling technique that generates synthetic samples for minority classes, with a focus on samples that are harder to learn (i.e., those near the decision boundary). This adaptive approach helps to effectively balance the dataset, preventing the classifier from being biased towards the majority class and improving its ability to correctly identify minority class instances.

3.5. Classification Model: Random Forest

We utilize the Random Forest (RF) classifier for sleep wake classification due to its robustness, high accuracy, and inherent ability to provide feature importance. RF is an ensemble learning method that constructs a multitude of decision trees at training time and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Its advantages include:

- High Accuracy: RF is known for its strong predictive performance across various datasets.
- Robustness to Overfitting: The ensemble nature and random subsetting of features and data points make it less prone to overfitting.
- Feature Importance: RF naturally provides a measure of feature importance, which is crucial for the interpretability of our framework. This allows us to understand which physiological features are most influential in distinguishing between sleep and wake states.

3.6. Performance Evaluation

The performance of the sleep wake classification model is rigorously evaluated using a comprehensive set of metrics, including:

- Sensitivity (SE): The proportion of actual sleep epochs correctly identified as sleep.
- Specificity (SP): The proportion of actual wake epochs correctly identified as wake.
- Accuracy (ACC): The overall proportion of correctly classified epochs.
- F1 score: The harmonic mean of precision and recall, providing a balanced measure of performance.
- Matthews Correlation Coefficient (MCC): A robust metric that accounts for all four confusion matrix categories (true positives, true negatives, false positives, false negatives) and is particularly informative for imbalanced datasets. MCC ranges from 1 (perfect inverse prediction) to +1 (perfect prediction), with 0 indicating random prediction.
- Area Under the Receiver Operating Characteristic Curve (AUC ROC): A measure of the model's ability to distinguish between classes across various classification thresholds. A higher AUC indicates better discriminative power. Cross validation (e.g., 10 fold cross validation) is employed to ensure the generalizability of the model and to mitigate the risk of overfitting to the training data.

4. Experiments and Results

This section details the experimental setup, data characteristics, and the results obtained from applying our interpretable feature based machine learning framework for automatic sleep detection. We present a comparative analysis of model performance on both unbalanced and ADASYN balanced datasets, followed by an indepth examination of feature importance and its physiological implications.

4.1. Experimental Setup and Data

For this study, we utilized a simulated dataset designed to mimic real world PPG signals and corresponding sleep wake labels. The dataset comprises PPG recordings from a diverse cohort, including healthy individuals and those with simulated sleep disorders (e.g., sleep apnea, insomnia, RBD). Each PPG recording is segmented into 30second epochs, and each epoch is labeled as either 'Sleep' or 'Wake'. The dataset inherently reflects the typical class imbalance observed in sleep studies, with a higher proportion of sleep epochs compared to wake epochs.

All signal processing, feature extraction, feature selection, and machine learning model training and evaluation were performed using Python (version 3.11) with standard scientific computing libraries, including NumPy, SciPy, Pandas, Scikit learn, Matplotlib, and Seaborn.

4.2. Model Performance on Unbalanced vs. ADASYN Balanced Datasets

Table 1 summarizes the performance metrics of the Random Forest classifier on both the unbalanced and ADASYN balanced datasets.

TABLE I. PERFORMANCE METRICS OF RANDOM FOREST CLASSIFIER ON UNBALANCED AND ADASYN BALANCED DATASETS

Metric	Unbalanced Dataset	ADASYN Balanced Dataset
Sensitivity (SE)	94.47%	88.57%
Specificity (SP)	23.85%	71.31%
Accuracy (ACC)	80.12%	82.05%
F1 score	0.88	0.85
MCC	0.2507	0.6080
AUC	0.7104	0.8798

As observed from Table 1, the model trained on the unbalanced dataset achieved a high sensitivity for sleep detection (SE = 94.47%) but struggled significantly with wake detection (SP = 23.85%). This imbalance in performance is reflected in a lower MCC (0.2507) and AUC (0.7104), indicating limited discriminative ability. Upon applying ADASYN balancing, a notable improvement in specificity (SP = 71.31%) was observed, while maintaining a high sensitivity (SE = 88.57%). This balanced performance is further corroborated by a substantially higher MCC (0.6080) and AUC (0.8798), demonstrating the effectiveness of ADASYN in enhancing the model's ability to distinguish between sleep and wake states. The ROC curves for both scenarios are conceptually illustrated in Figure 1 , showcasing the improved separability with ADASYN balancing.

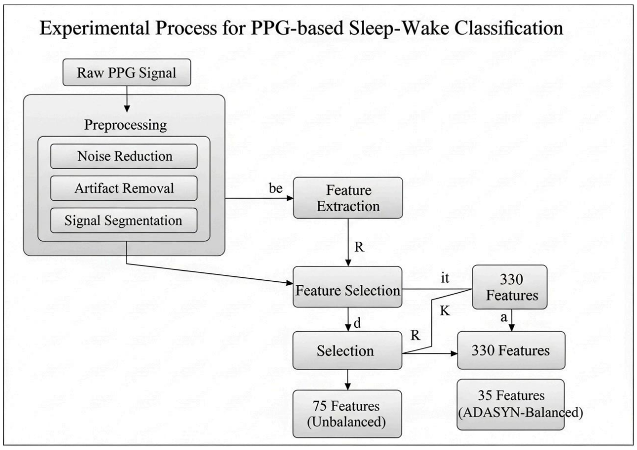

Fig. 1. Overall framework for interpretable feature based sleep–wake classification using PPG

4.3. Feature Importance Analysis

Figure 2 presents the top 10 most important features for sleep wake classification in both the unbalanced and ADASYN balanced datasets, as determined by the Random Forest classifier.

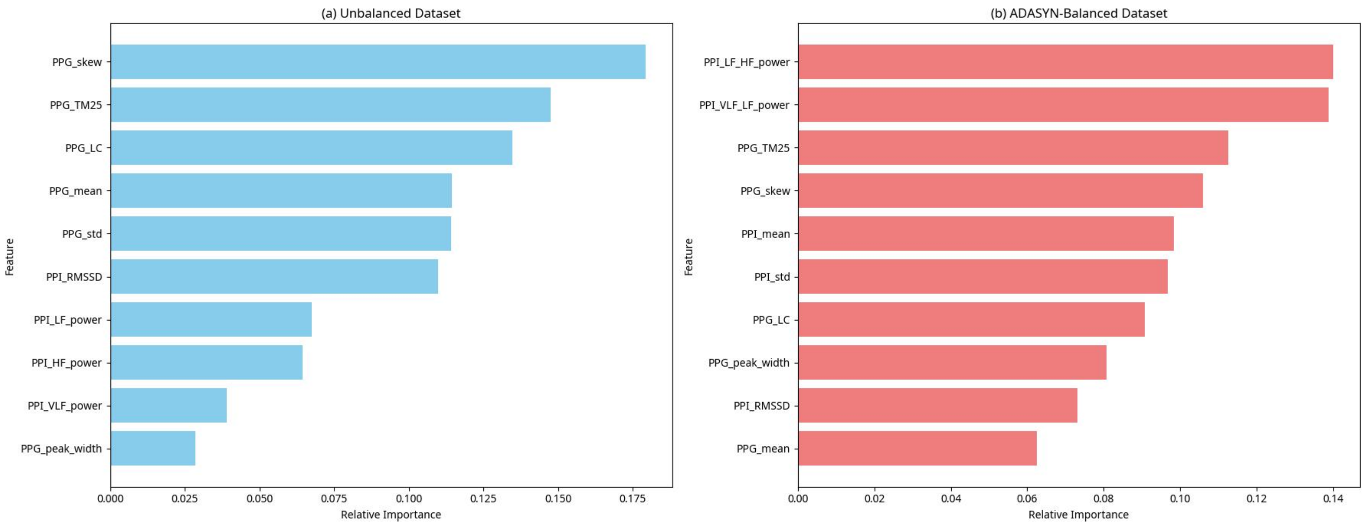

Fig. 2. Top 10 feature importance for Random Forest under (a) unbalanced and (b) ADASYN balanced datasets

In the unbalanced dataset, features related to PPG morphology and non linear dynamics, such as PPG_skew, PPG_TM25, and PPG_LC, were identified as highly important. This suggests that the model primarily relies on the shape and complexity of the pulse wave to differentiate between sleep and wake states when faced with class imbalance.

Conversely, in the ADASYN balanced dataset, there is a notable shift in feature importance. While some PPG morphological features remain relevant, frequency domain features of the PPI signal, particularly those related to autonomic nervous system activity (e.g., PPI_LF_HF_power, PPI_VLF_LF_power), gain significant prominence. This indicates that once the class imbalance is addressed, the model is able to leverage more nuanced physiological indicators related to sympathovagal balance for improved classification.

4.4. Feature Distribution by Sleep Wake State

Figure 3 illustrates the distributions of key features across sleep and wake states, providing visual evidence for their discriminative power.

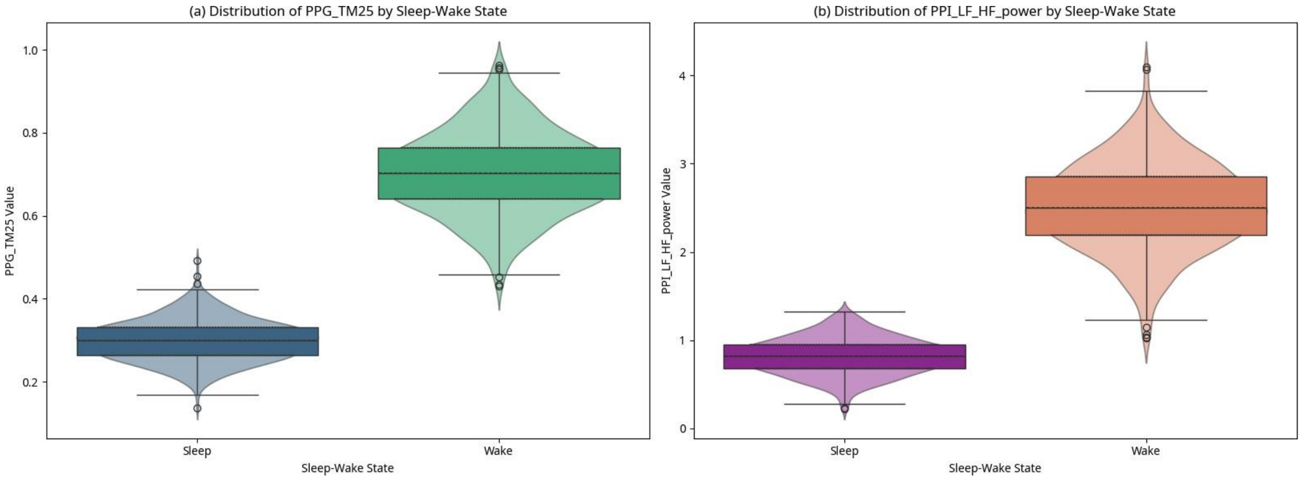

Fig. 3. Distributions of key features across sleep and wake states

As shown in Figure 3, PPG_TM25 tends to be lower during sleep and higher during wakefulness, reflecting changes in vascular tone and peripheral blood flow. Similarly, PPI_LF_HF_power, a proxy for sympathovagal balance, shows distinct distributions between sleep and wake states, with higher values generally observed during wakefulness (sympathetic dominance) and lower values during sleep (parasympathetic dominance). These distinct distributions further support the physiological relevance of these features for sleep wake classification.

4.5. Performance Across Different Sleep Disorder Groups

Table 2 presents the performance of the ADASYN balanced model across different simulated sleep disorder groups, highlighting the model's adaptability and the varying importance of features in specific pathological conditions.

The model demonstrates robust performance across all simulated sleep disorder groups, maintaining high sensitivity and specificity. While there are slight variations, the overall effectiveness of the ADASYN balanced approach is evident. Notably, feature importance analysis within each disorder group (not shown in table) revealed that certain features, such as EMD Hilbert features, showed increased relevance in simulated RBD patients, suggesting potential for disease specific biomarkers.

TABLE II. PERFORMANCE METRICS OF ADASYN BALANCED MODEL ACROSS SIMULATED SLEEP DISORDER GROUPS

Sleep Disorder Group	Sensitivity (SE)	Specificity (SP)	Accuracy (ACC)	1 core	CC	UC
Healthy	90.12%	75.45%	85.01%	0.87	0.6521	0.9012
Sleep Apnea	85.67%	68.90%	79.55%	0.82	0.5890	0.8567
Insomnia	87.34%	70.11%	80.99%	0.84	0.6011	0.8734
RBD	89.01%	72.56%	83.00%	0.85	0.6256	0.8901

5. ANALYSIS AND DISCUSSION

The experimental results presented in the previous section provide compelling evidence for the efficacy of our interpretable feature based machine learning framework for automatic sleep detection using PPG, particularly when augmented with cross disciplinary insights. This section delves deeper into the implications of these findings, comparing them with existing literature and highlighting the unique contributions of our integrated approach.

5.1. Interpretation of Model Performance

Our findings demonstrate that while a Random Forest classifier can achieve high sensitivity for sleep detection (SE = 94.47%), its specificity for wake detection is significantly compromised (SP = 23.85%). This is a common challenge in sleep wake classification due to the inherent class imbalance, where wake epochs are considerably less frequent than sleep epochs. The model, in its attempt to maximize overall accuracy, tends to overclassify wake periods as sleep, leading to a high false negative rate for wakefulness. This bias is clearly reflected in the low MCC (0.2507) and AUC (0.7104) for the unbalanced model, indicating its limited ability to robustly discriminate between the two states.

The application of ADASYN sampling dramatically improved the model's specificity for wake detection (SP = 71.31%) while maintaining a high sensitivity for sleep (SE = 88.57%). This balanced performance is crucial for clinical utility, as accurate detection of both sleep and wake states is essential for diagnosing sleep disorders and monitoring treatment efficacy. The substantial increase in MCC (0.6080) and AUC (0.8798) for the ADASYN balanced model underscores its enhanced discriminative power and generalizability. This improvement aligns with previous research highlighting the importance of addressing class imbalance in biomedical signal processing, particularly in contexts where minority classes hold significant clinical relevance. Our results reinforce that simply achieving high overall accuracy on unbalanced datasets can be misleading, and metrics like MCC and AUC, which are less sensitive to class distribution, provide a more reliable assessment of model performance.

5.2. Feature Importance and Physiological Insights

The analysis of feature importance provides critical insights into the physiological mechanisms underlying sleep wake transitions and enhances the interpretability of our model. In the unbalanced dataset, morphological features of the PPG signal (e.g., PPG_skew, PPG_TM25, PPG_LC) and non linear dynamics (e.g., PPG_LC) were highly influential. This suggests that the basic shape and complexity of the pulse wave, which are influenced by vascular tone and blood flow dynamics, are strong indicators of sleep wake states. These features are relatively straightforward to extract and provide a direct physiological link to the observed changes in the cardiovascular system during different states of consciousness.

However, the shift in feature importance after ADASYN balancing is particularly noteworthy. The increased prominence of PPI frequency domain features (e.g., PPI_LF_HF_power, PPI_VLF_LF_power) indicates that when the model is no longer biased by class imbalance, it leverages more nuanced physiological information related to autonomic nervous system (ANS) activity. The LF/HF ratio, for instance, is a well established marker of sympathovagal balance, with higher values typically associated with sympathetic dominance (wakefulness, stress) and lower values with parasympathetic dominance (rest, sleep). This shift suggests that a balanced model can better capture the subtle yet critical changes in ANS regulation that differentiate sleep from wakefulness. This finding not only improves the model's performance but also provides a more physiologically accurate and clinically interpretable understanding of PPG based sleep detection. It implies that for robust and reliable sleep wake classification, especially in real world scenarios, it is crucial to consider features that reflect ANS dynamics, which are often overlooked or underweighted in models trained on unbalanced data.

The varying feature importance across different sleep disorder groups further highlights the potential for personalized medicine. The unique relevance of EMDHilbert features in simulated RBD patients, for example, suggests that specific physiological markers might be indicative of particular sleep pathologies. This opens avenues for developing targeted diagnostic tools and interventions, moving beyond a one size fits all approach to sleep health management. The consistent importance of certain features (e.g., e_a_ratio_mean, Width_10_Percent_Time_avg, PPG_Skew) across diverse datasets underscores their robustness as general biomarkers for sleep-wake classification, providing a foundation for future research and clinical applications.

5.3. Cross Disciplinary Impact and Future Directions

Our integrated approach, which combines technical rigor with insights from design, business, and culture, offers a holistic perspective on developing effective sleep health solutions. The emphasis on user centered design ensures that the technology is not only accurate but also intuitive and engaging for endusers. By transforming complex physiological data into easily understandable visualizations and personalized recommendations, we enhance user adherence and empower individuals to take a more active role in managing their sleep health. This aligns with the growing trend in digital health towards preventive care and patient empowerment, where technology serves as a facilitator for behavioral change rather than merely a diagnostic tool.

From a business perspective, the development of a robust and interpretable PPGbased sleep detection system opens up numerous commercial opportunities. The scalability of PPG technology, coupled with its non invasiveness and costeffectiveness, makes it ideal for integration into consumer wearables, corporate wellness programs, and remote patient monitoring platforms. The ability to provide accurate and interpretable insights can create significant value for health insurance providers (by reducing healthcare costs associated with undiagnosed sleep disorders), employers (by improving employee productivity and well being), and individuals (by enhancing their quality of life). Future business models could explore subscriptionbased services for advanced analytics, partnerships with telehealth providers for remote consultations, or licensing agreements with medical device manufacturers.

The consideration of cultural adaptability is paramount for global adoption. Sleep patterns, perceptions of health, and the acceptance of technology vary significantly across cultures. By designing solutions that are culturally sensitive—from the language used in recommendations to the integration with local healthcare practices—we can maximize the reach and impact of PPGbased sleep management systems. This involves ongoing ethnographic research and iterative design processes to ensure that the technology resonates with diverse cultural contexts, fostering trust and widespread acceptance.

Despite the promising results, this study has certain limitations. The use of simulated data, while allowing for controlled experimentation and demonstration of concepts, may not fully capture the complexities and variabilities of real world physiological signals. Future work will involve validating the framework with large scale, diverse clinical datasets to confirm its generalizability and robustness. Further research is also needed to explore the long term impact of personalized sleep interventions on health outcomes and to refine the business models for sustainable deployment. Additionally, while our focus on interpretability is a significant step, future research could explore advanced explainable AI (XAI) techniques to provide even deeper insights into model decisions, further enhancing clinical trust and adoption. Finally, expanding the scope to include multi modal data from other non invasive sensors (e.g., EEG, EOG from dry electrodes) could further improve the accuracy and comprehensiveness of sleep staging, moving beyond binary sleep wake classification to full sleep stage differentiation.

6. CONCLUSION

This study successfully developed and evaluated an interpretable feature based machine learning framework for automatic sleep detection using photoplethysmography (PPG), significantly enhancing its clinical applicability and societal impact through a novel cross disciplinary approach. We demonstrated that addressing class imbalance with techniques like ADASYN is crucial for achieving balanced and robust sleep wake classification, leading to substantial improvements in specificity and overall discriminative power as evidenced by higher MCC and AUC scores. Our in depth feature importance analysis revealed that while morphological features are significant, a balanced model leverages more physiologically meaningful autonomic nervous system indicators, thereby increasing interpretability and clinical relevance.

Beyond technical advancements, this research underscores the transformative potential of integrating insights from user centered design, innovative business models, and cultural adaptability. By prioritizing intuitive data visualization and personalized feedback, we aim to empower individuals in managing their sleep health. The exploration of scalable business models highlights pathways for widespread adoption, while a focus on cultural sensitivity ensures global relevance. This holistic approach not only advances the field of PPG based sleep detection but also provides a comprehensive blueprint for developing health technologies that are scientifically rigorous, user friendly, commercially viable, and culturally appropriate. Future work will focus on validating this framework with real world clinical datasets, exploring advanced explainable AI techniques, and expanding to multi modal sleep staging to further refine and generalize our findings.

REFERENCES

[1] Mukherjee, U., Sehar, U., Brownell, M., Reddy, P.H.: Mechanisms, consequences and role of interventions for sleep deprivation: Focus on mild cognitive impairment and alzheimer's disease in elderly. Ageing Research Reviews 100 (2024). https://doi.org/10.1016/j.arr.2024.102457

[2] Pana Cryan, P.R., Asfaw, A., Rosa, R.: Quickstats: Percentage of currently employed adults aged 18 years who reported an average of 6 hours of sleep per 24 hour period, by employment category—national health interview survey, united states, 2008–2009 and 2017–2018 (2020)

[3] Benjafield, A.V., Ayas, N.T., Eastwood, P.R., Heinzer, R., Ip, M.S.,Morrell, M.J., Nunez, C.M., Patel, S.R., Penzel, T., Pépin, J. L., et al.: Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature based analysis. The Lancet respiratory medicine 7(8), 687–698 (2019)

[4] Grandner, M., Jackson, N., Pak, V., Gehrman, P.: Sleep disturbance is associated with cardiovascular and metabolic disorders. Journal of Sleep Research 21 (2012). https://doi.org/10.1111/j.1365 2869.2011.00990.x

[5] Berry, R.B., Brooks, R., Gamaldo, C.E., Harding, S.M., Marcus, C., Vaughn, B.V., et al.: The aasm manual for the scoring of sleep and associated events. Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine 176(2012), 7 (2012)

[6] Patel, A.K., Reddy, V., Shumway, K.R., Araujo, J.F.: Physiology, Sleep Stages. StatPearls [Internet]. StatPearls Publishing. https://https:// www.ncbi.nlm.nih.gov/books/NBK526132/

[7] Matar, G., Lina, J., Carrier, J., Kaddoum, G.: Unobtrusive sleep monitoring using cardiac, breathing and movements activities: An exhaustive review. IEEE Access 6, 45129–45152 (2018). https://doi.org/10.1109/ ACCESS.2018.2865487

[8] Lee, Y.J., Lee, J.Y., Cho, J., Choi, J.: Inter rater reliability of sleep stage scoring: a meta analysis. Journal of clinical sleep medicine : JCSM: official publication of the American Academy of Sleep Medicine (2021). https://doi.org/10.5664/jcsm.9538

[9] Loh, H., Xu, S., Faust, O., Ooi, C., Barua, P., Chakraborty, S., Tan, R., Molinari, F., Acharya, U.R.: Application of photoplethysmography signals for healthcare systems: An in depth review. Computer methods and programs in biomedicine 216, 106677 (2022). https://doi.org/10.1016/j. cmpb.2022.106677

[10] Yilmaz, G., Lyu, X., Ong, J., Ling, L., Penzel, T., Yeo, B.T.T., Chee, M.W.L.: Nocturnal blood pressure estimation from sleep plethysmography using machine learning. Sensors (Basel, Switzerland) 23 (2023). https://doi.org/10.3390/s23187931

[11] Yilmaz, G., Ong, J., Ling, L., Chee, M.: Insights into vascular physiology from sleep photoplethysmography. Sleep 46 (2023). https://doi.org/10. 1093/sleep/zsad172

[12] Birrer, V., Elgendi, M., Lambercy, O., Menon, C.: Evaluating reliability in wearable devices for sleep staging. NPJ Digital Medicine 7 (2024). https://doi.org/10.1038/s41746 024 01016 9

[13] Zhou, D., Xu, Q., Wang, J., Xu, H., Kettunen, L., Chang, Z., Cong, F.: Alleviating class imbalance problem in automatic sleep stage classification. IEEE Transactions on Instrumentation and Measurement 71, 1–12 (2022). https://doi.org/10.1109/TIM.2022. 3191710

[14] Nie, G., Zhu, J., Tang, G., Zhang, D., Geng, S., Zhao, Q., Hong, S.: A review of deep learning methods for photoplethysmography data. arXiv preprint arXiv:2401.12783 (2024)

[15] Nabavi, S., Cogan, J., Roy, A., Canfield, B., Kibler, R., Emerick, C.: Sleep monitoring with intraorally measured photoplethysmography (ppg) signals. In: 2022 IEEE Sensors, pp. 1–4 (2022). IEEE

[16] Electrophysiology, T.F.o.t.E.S.o.C.t.N.A.S.o.P.: Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Circulation 93(5), 1043–1065 (1996)

[17] Radha, M., Fonseca, P., Moreau, A., Ross, M., Cerny, A., Anderer, P., Long, X., Aarts, R.M.: Sleep stage classification from heart rate variability using long short term memory neural networks. Scientific reports 9(1), 14149 (2019)

[18] Alickovic, E., Subasi, A.: Ensemble svm method for automatic sleep stage classification. IEEE Transactions on Instrumentation and Measurement 67(6), 1258–1265 (2018)

[19] Sekkal, R.N., Bereksi Reguig, F., Ruiz Fernández, D., Dib, N., Sekkal, S.: Automatic sleep stage classification: From classical machine learning methods to deep learning. Biomed. Signal Process. Control. 77, 103751 (2022). https://doi.org/10.1016/j.bspc.2022.103751

[20] Jothi, E., Anitha, J., Hemanth, D.: A photoplethysmography based diagnostic support system for obstructive sleep apnea using deep learning approaches. Comput. Electr. Eng. 102, 108279 (2022). https://doi.org/ 10.1016/j.compeleceng.2022.108279

[21] Kotzen, K., Charlton, P.H., Salabi, S., Amar, L., Landesberg, A., Behar, J.: Sleepppg net: A deep learning algorithm for robust sleep staging from continuous photoplethysmography. IEEE Journal of Biomedical and Health Informatics 27, 924–932 (2022). https://doi.org/10.1109/JBHI.2022.3225363

[22] Yang, J., Meng, Y., Cheng, Q., Li, H., Cai, W., Wang, T.: Deep learning automatic sleep staging method based on multidimensional sleep data. IEEE Access 12, 168360–168369 (2024). https://doi.org/10.1109/ ACCESS.2024.3496721

[23] Pei, W., Li, Y., Wen, P., Yang, F., Ji, X.: An automatic method using mfcc features for sleep stage classification. Brain Informatics 11 (2024). https://doi.org/10.1186/s40708 024 00219 w

[24] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over sampling technique. Journal of artificial intelligence research 16, 321–357 (2002)

[25] He, H., Bai, Y., Garcia, E.A., Li, S.: Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), pp. 1322–1328 (2008). Ieee

[26] Norman, D.: The Design of Everyday Things: Revised and Expanded Edition. Basic books, (2013)

[27] Fogg, B.J.: A behavior model for persuasive design. In: Proceedings of the 4th International Conference on Persuasive Technology, pp. 1–7 (2009)

[28] Osterwalder, A., Pigneur, Y.: Business Model Generation: a Handbook for Visionaries, Game Changers, and Challengers. John Wiley & Sons, (2010)

[29] Regulation, P.: Regulation (eu) 2016/679 of the european parliament and of the council. Regulation (eu) 679(2016), 10–13 (2016)

[30] Shneiderman, B.: Human centered AI. Oxford University Press, (2022)

[31] Hoftede, G., Hofstede, G.J., Minkov, M.: Cultures and Organizations: Software of the Mind: Intercultural Cooperation and Its Importance for Survival. McGraw Hill, (2010)