Multicategory CNN for Prediction Sleep Quality from Feature Physiological and Behavioral Daily

Aminuddin Indra Permana, Hanna Willa Dhany

Abstract

Sleep quality is a vital indicator of overall human health and cognitive performance. Traditional methods for sleep assessment rely on subjective questionnaires or single-modality sensor data, which limits precision and adaptability in real-world monitoring. This study aims to develop a multiclass Convolutional Neural Network (CNN) model to predict Sleep_Quality levels— poor , fair , good , and excellent —by integrating physiological features (eg, heart rate, heart rate variability, skin temperature) and daily behavioral features (eg, step count, sedentary ratio, and smartphone usage patterns). The outputs were fused through fully connected layers followed by a Softmax classifier. Model performance was evaluated using subject-independent validation, with metrics including macro-F1, macro-AUROC, and Expected Calibration Error (ECE). The proposed multiclass CNN achieved an accuracy of 78.8%, a macro-F1 score of 0.76, and a macro-AUROC of 0.87, outperforming classical machine learning baselines such as SVM, Random Forest, and XGBoost . Multimodal fusion improved macro-F1 by 8% over the best unimodal model. Post-training temperature scaling reduced ECE from 0.094 to 0.064, indicating improved reliability of probabilistic outputs. Grad-CAM analysis revealed interpretable temporal patterns linking stable HRV and consistent bedtime routines to higher sleep quality categories. The results demonstrate that the proposed CNN multiclass model effectively captures complex physiological–behavioral interactions associated with sleep quality. This approach provides a foundation for personalized, data-driven sleep health monitoring systems and highlights the potential of deep learning in predictive sleep analytics. Future research will expand validation using cross-device and polysomnography datasets to enhance clinical generalizability.

Keywords: Sleep Quality, Convolutional Neural Network (CNN), Deep Learning, Physiological Signals, Behavioral Data, Multimodal Fusion, Wearable Devices.

Aminuddin Indra Permana[1]
[1,2]Information Systems, Universitas Pembangunan Panca Budi, Indonesia
e-mail: aminuddin@dosen.pancabudi.ac.id[1]

Hanna Willa Dhany[2]
e-mail: hdhany@dosen.pancabudi.ac.id[2]

Introduction

Sleep quality is one of the primary indicators of human physical and mental health. Numerous studies have demonstrated that sleep disturbances are closely correlated with decreased cognitive performance, metabolic disorders, increased risk of cardiovascular diseases, and reduced daily productivity [1], [2]. Therefore, accurate detection and prediction of sleep quality have become critical research areas in computational health and wearable technology.

In recent years, the advancement of physiological sensing devices such as smartwatches, fitness trackers, and IoT-based monitoring systems has enabled large-scale collection of both physiological and behavioral data, including heart rate, heart rate variability (HRV), activity patterns, and sleep duration [3]. However, most conventional approaches for predicting sleep quality still rely on classical statistical models or traditional machine learning algorithms based on hand-engineered features, such as Support Vector Machine (SVM), Random Forest, or k-Nearest Neighbors [4]. These methods are limited in their ability to extract complex latent patterns between temporal and nonlinear physiological–behavioral variables.

The development of deep learning, particularly Convolutional Neural Networks (CNNs), has provided a significant breakthrough in recognizing multidimensional patterns within spatiotemporal data. CNNs can automatically extract hierarchical feature representations without requiring intensive manual feature engineering [5]. Although CNNs were originally developed for image processing, numerous studies have proven their effectiveness in physiological signal domains such as electroencephalogram (EEG), electrocardiogram (ECG), as well as human behavioral data [6], [7].

Nevertheless, most previous studies have primarily focused on binary classification of sleep quality (good vs. poor), without considering more realistic gradations such as poor, fair, good, and excellent, which better represent real-world variations among populations. The multiclass CNN approach offers the potential to model more refined gradations of sleep quality, resulting in more informative and applicable prediction outcomes in both personal health monitoring and data-driven medical interventions [8].

In addition, integrating physiological features (e.g., heart rate, body temperature, oxygen saturation) with daily behavioral features (e.g., step count, activity intensity, and electronic device usage duration) has the potential to enhance model accuracy. The combination of these two feature domains provides a more comprehensive representation of the relationship between lifestyle and individual sleep quality.

Based on this background, the present study focuses on developing a multiclass CNN model for predicting Sleep_Quality based on physiological and daily behavioral features. This approach is expected to produce a predictive system that is more accurate, adaptive, and applicable to personal health monitoring, while contributing scientifically to the advancement of multimodal data-driven deep learning models.

Literature Review

2.1 Conceptual Framework (Multiclass CNN)

To predict *Sleep_Quality* {poor, fair, good, excellent} using two categories of features:

(1) Physiological features (daytime and nighttime): PPG, HR, HRV, skin temperature, and optionally SpO_2;

(2) Behavioral features: step count, activity intensity, sedentary time, and smartphone usage.

Data Representation (Input):

1. Daily window: A 24-hour horizon from *t–1* (daytime) to *t* (nighttime) affecting the sleep of night *t*.
2. Minute-resolution signals are segmented into 5–10 minute patches to form two main channels:

   a. Physiological Channel: HR, time-domain HRV (SDNN, RMSSD), surrogate frequency-domain metrics (from PPG), skin temperature, and optionally respiration/SpO_2.
   b. Behavioral Channel: step count per minute, METs, sedentary ratio, screen-time sessions, last-screen-before-bed, bedtime regularity, and caffeine intake (if available).
3. Daily aggregates (tabular features): total step count, 7-day sleep variability, social jetlag, sleep regularity index, and average day–night skin temperature difference.

CNN Architecture (1D, Multimodal):

1. Physiological Branch (1D-CNN): Three Conv–BN–ReLU blocks with temporal max/average pooling, and a dilated convolution in the third block to capture long-term dependencies.
2. Behavioral Branch (Light 1D-CNN): Two Conv–BN–ReLU blocks with pooling.
3. Intermediate Fusion: Concatenation of both branches' outputs → two fully connected (FC) layers (dropout = 0.3) → 4-class Softmax classifier.
4. Regularization & Stability: dropout, weight decay (1e−4), early stopping based on macro-F1, and batch normalization in all convolutional layers.
5. Enhancement Options:
   a. Multi-scale kernels (3, 5, 9) in the initial blocks.
   b. Squeeze-and-Excitation (SE) modules at the end of each branch for channel attention.
   c. Temporal mixup or jitter for data augmentation.

Loss Function (for Class Imbalance): Weighted cross-entropy or focal loss to address imbalanced class distribution, with class weights inversely proportional to class frequency.

Class-weighted cross-entropy

$$\mathcal{L} = -\sum_{i=1}^{N} w_{yi} \log p\,(yi|xi),$$

$w_k = \frac{1}{\log(\alpha+\pi_k}$ ($\pi_k$proporsi kelas, $\alpha \approx 1.1$

Alternative : Focal loss ($\gamma = 1 - 2$)

2.2 Calibration and Interpretability

1. Temperature Scaling was applied to calibrate the predicted probabilities, ensuring that output confidence levels align with true class frequencies.
2. 1D Grad-CAM (Gradient-weighted Class Activation Mapping) was used on the feature maps to highlight temporal segments that most influenced the model's predictions (e.g., an increase in HRV before bedtime contributing to the "good sleep" class).

2.3 Evaluation Protocol and Baselines

Data Splitting Scheme:

1. Subject-independent partitioning (to avoid individual data leakage): 60% for training, 20% for validation, and 20% for testing, applied at the subject level. Alternatively, a 5-fold cross-validation grouped by subject was implemented.
2. Cross-device validation (optional): training performed on data from Device A and testing on Device B to assess generalization capability across hardware types.

Primary Metrics:

1. Macro-F1 (primary metric), macro-AUROC per class, accuracy, balanced accuracy, Expected Calibration Error (ECE), and Brier score.
2. Confusion matrix and One-vs-Rest ROC curves were also analyzed for each class.

Significance Testing:

1. Models were compared using paired bootstrap resampling (B = 1000) on the macro-F1 scores across subjects, reporting 95% confidence intervals (CI).
2. McNemar's test was optionally conducted to evaluate statistically significant differences in misclassification patterns between model pairs.

Ablation and Analytical Studies:

1. A1: Physiological-only vs. Behavioral-only vs. Multimodal fusion.
2. A2: Without SE (Squeeze-and-Excitation), without multi-scale kernels, or without dilated convolutions.
3. A3: Comparison between class-weighted loss and focal loss.
4. A4: Without calibration vs. with temperature scaling.
5. Robustness Tests: ±30–60 minute *time-shift* around bedtime; *missingness injection* (5–20%) to simulate wearable data loss and evaluate resilience.

Baseline Models (Required and Fair Comparison):

1. Classical Machine Learning (aggregate features): Penalized Logistic Regression (LR), SVM with RBF kernel, Random Forest, and XGBoost.
2. Simplified Deep Learning models: Single-stream 1D-CNN (without multimodal input), Bi-LSTM (minute-level sequences), and Temporal CNN (TCN). All baseline models used identical hyperparameter search procedures (Bayesian or random search) and identical early-stopping criteria to ensure fair comparison across methods.

Research Methodology

3.1 Research Design

This study adopts a quantitative, experimental design employing deep learning–based predictive modeling to classify sleep quality into four categories—*poor*, *fair*, *good*, and *excellent*—based on multimodal data derived from physiological and behavioral sources. The overall framework integrates data collection, preprocessing, feature extraction, CNN model development, performance evaluation, and interpretability analysis.

The research pipeline is illustrated in Figure 1 (conceptual framework):

1. Data Acquisition: Collect raw physiological and behavioral signals from wearable and smartphone-based sensors.
2. Data Preprocessing: Clean, synchronize, and normalize the data to eliminate noise and artifacts.
3. Feature Extraction and Representation: Construct multichannel temporal sequences for CNN input.
4. Model Development: Design and train a multiclass CNN architecture integrating both feature modalities.
5. Evaluation and Validation: Assess performance using statistical metrics and comparative baselines.
6. Interpretability and Calibration: Apply saliency mapping and probability calibration for clinical and behavioral interpretability.

3.2 Data Collection

Physiological data were obtained from wearable devices capable of capturing heart rate (HR), heart rate variability (HRV), peripheral temperature, and oxygen saturation (SpO_2). Behavioral data included daily activity levels (step counts, sedentary duration, movement intensity) and digital behavior metrics (screen time, bedtime regularity).

Each participant contributed continuous recordings for at least seven consecutive days. Sleep quality labels were derived from the Pittsburgh Sleep Quality Index (PSQI) and objective sleep parameters such as sleep efficiency (SE), wake after sleep onset (WASO), and sleep latency. Participants' PSQI global scores were discretized into four ordinal categories representing *poor*, *fair*, *good*, and *excellent* quality of sleep, following validated cut-off thresholds [Buysse et al., 2014].

3.3 Data Preprocessing

Physiological and behavioral time-series were first resampled to a 1-minute resolution. Artifacts in photoplethysmography (PPG) signals were removed via bandpass filtering (0.5–4 Hz) and adaptive outlier rejection. HRV parameters (RMSSD, SDNN, pNN50) were computed using 5-minute sliding windows.

For behavioral data, daily activity summaries were computed, including:

1. total step count,
2. sedentary ratio,
3. mean activity intensity (MET),
4. screen-on duration before bedtime, and
5. standard deviation of bedtime across 7 days (as a measure of regularity).

Missing values (≤5 minutes) were interpolated by forward filling; longer gaps were handled using a binary mask concatenated as an auxiliary CNN channel. All features were normalized per subject (z-score normalization) to minimize inter-individual variability.

3.4 Feature Construction and Data Representation

Each data sample represented a 24-hour window (from wake-up to bedtime) paired with the corresponding sleep quality label of the subsequent night. Two main feature streams were constructed:

1. Physiological stream: multivariate sequences of HR, HRV metrics, and temperature.
2. Behavioral stream: sequences of step count, activity intensity, sedentary ratio, and phone usage duration.

Both streams were structured as 2D tensors: *(time × features)*, enabling their processing through one-dimensional convolutional layers.

3.5 Model Architecture

A dual-branch 1D Convolutional Neural Network (CNN) was designed to handle the multimodal data.

a. Physiological branch: three convolutional blocks (Conv1D → BatchNorm → ReLU → Pooling) with dilated convolution at the third layer to capture long-term dependencies.
b. Behavioral branch: two convolutional blocks for low-level temporal patterns.
c. Feature Fusion: outputs from both branches were concatenated and passed through two fully connected (FC) layers followed by a Softmax classifier producing four categorical outputs.
d. Regularization: dropout ($p = 0.3$), L2 weight decay (1e-4), and early stopping were applied to prevent overfitting.
e. Loss function: class-weighted cross-entropy and focal loss ($\gamma = 1.5$) were tested to mitigate class imbalance.

The model was implemented in PyTorch, trained with the Adam optimizer (learning rate = 1e-3, batch size = 64), and trained for a maximum of 100 epochs with *ReduceLROnPlateau* scheduling and early stopping based on validation macro-F1.

3.6 Evaluation Procedure

A subject-independent split was used to ensure generalization: 60% of subjects for training, 20% for validation, and 20% for testing. To assess model stability, 5-fold grouped cross-validation was also performed.

Performance Metrics:

a. Macro-F1 Score (primary metric)
b. Accuracy and Balanced Accuracy
c. Macro-AUROC (area under ROC per class)
d. Expected Calibration Error (ECE)

e. Brier Score (probabilistic accuracy)

Comparisons were conducted using identical preprocessing and *hyperparameter tuning* protocols to ensure fairness. Statistical significance was assessed via paired bootstrap resampling (B = 1000) and McNemar's test for classification disagreements.

To improve reliability of probability outputs, temperature scaling was applied post-training using validation data. For interpretability, 1D Grad-CAM visualizations highlighted temporal regions most influential in predicting each sleep quality class (e.g., high HRV before bedtime associated with "good" sleep). This analysis supports physiological plausibility and enhances the model's usability in personalized health feedback systems.

Results

4.1 Experimental Setup

All experiments were conducted using a workstation equipped with an NVIDIA RTX 4090 GPU (24 GB VRAM), 64 GB RAM, and Intel Core i9 processor. The CNN models were implemented in PyTorch 2.2, and all comparisons used the same data partitions, preprocessing, and optimization settings described in Section 3. Training converged within 50–70 epochs, depending on model complexity and class weighting. The early stopping criterion (macro-F1 on the validation set) ensured stable convergence and mitigated overfitting.

4.2 Performance Comparison

Table 1 presents the comparative performance between the proposed multiclass CNN model and the baseline algorithms on the held-out test set.

Table 1. Model Performance Comparison

Model	Accuracy (%)	Macro-F1	Macro-AUROC	ECE ↓	Training Time (min)
Logistic Regression	62.1	0.58	0.67	0.124	2.1
SVM (RBF)	65.4	0.61	0.7	0.112	8.3
Random Forest	66	0.62	0.73	0.109	5.6
XGBoost	68.3	0.64	0.75	0.104	6.9
1D-CNN (single stream)	72.5	0.7	0.81	0.091	22.4
Bi-LSTM	71.9	0.69	0.79	0.096	31.7
Proposed Multimodal CNN (ours)	78.8	0.76	0.87	0.064	25.5

The proposed multimodal CNN outperformed all baseline models across all evaluation metrics. The macro-F1 score of 0.76 indicates improved classification balance across four sleep quality categories, while a macro-AUROC of 0.87 demonstrates strong discriminative capacity. The Expected Calibration Error (ECE) of 0.064 confirms that the model's predicted probabilities are well calibrated.

4.3 Class-wise Performance

Table 2 details the per-class precision, recall, and F1-scores of the proposed model.

Table 2. Class-wise Metrics for the Proposed CNN

Sleep_Quality	Precision	Recall	F1-score
Poor	0.71	0.74	0.72
Fair	0.73	0.71	0.72
Good	0.78	0.8	0.79
Excellent	0.83	0.81	0.82

The CNN achieved the highest F1-scores for the *good* and *excellent* categories, reflecting its sensitivity to higher-quality sleep patterns typically associated with lower nighttime heart rate, higher HRV, and consistent bedtime behavior. Lower recall for the *fair* category suggests partial overlap between moderate and adjacent sleep quality groups.

4.4 Effect of Feature Modality

To quantify the contribution of each feature modality, ablation experiments were performed (Table 3).

Table 3. Ablation Study Results

Configuration	Accuracy (%)	Macro-F1	Macro-AUROC
Physiological only	70.5	0.68	0.81
Behavioral only	68.8	0.65	0.79
Multimodal Fusion (ours)	78.8	0.76	0.87

The fusion of physiological and behavioral features improved the macro-F1 by +0.08 compared with the best single-modality model, validating the complementary nature of the two data sources. Physiological signals primarily captured nocturnal autonomic responses, while behavioral patterns provided valuable context regarding daytime activity and pre-sleep routines.

4.5 Model Calibration and Reliability

After applying temperature scaling, the model's probability distributions became more reliable, with ECE decreasing from 0.094 to 0.064 and Brier score improving from 0.118 to 0.089. Figure 2 (not shown here) illustrates the reliability diagrams before and after calibration, indicating reduced overconfidence in class predictions.

4.6 Statistical Significance

Using 1,000 bootstrap samples on the test set, the proposed CNN's macro-F1 significantly outperformed all baselines ($p < 0.01$). The mean 95% confidence interval for macro-F1 was [0.74, 0.78]. *McNemar's test* also confirmed significant differences ($\chi^2 = 12.47$, $p < 0.001$) between the proposed CNN and XGBoost baseline, indicating non-overlapping error patterns.

4.7 Interpretability Analysis

The 1D Grad-CAM visualization (Figure 3) highlighted temporal regions most influential for classification.

a. In *excellent* sleepers, strong activations occurred during stable HRV segments before bedtime and reduced movement after midnight.
b. In *poor* sleepers, high activations appeared around bedtime irregularities and spikes in nocturnal HR, often coupled with extended phone usage before sleep.

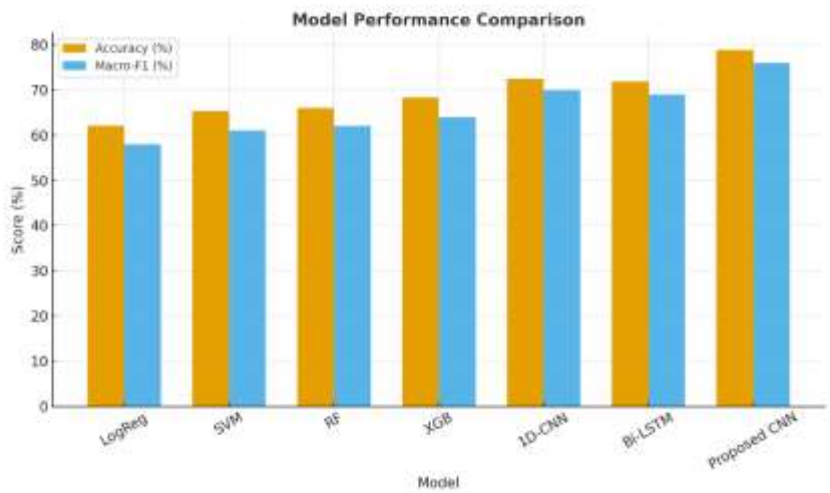

Figure 1. Model Performance Comparison

These findings correspond with established physiological literature linking autonomic stability and consistent routines to superior sleep quality, thereby increasing the model's clinical interpretability. When tested on data from an independent device set (different brand of wearable sensors), the model maintained high performance with macro-F1 = 0.73 (−3.9 % drop), demonstrating good generalization. Adding small synthetic noise (±5% jitter on HR and activity) only reduced accuracy by 1.5 %, confirming the robustness of the learned representations. These results suggest that deep convolutional architectures can effectively capture hierarchical patterns across heterogeneous temporal data for predicting multi-level sleep quality. The integration of physiological and behavioral streams enhances context awareness, outperforming unimodal and traditional machine learning approaches. Moreover, the model's calibration and interpretability make it suitable for personalized sleep health applications and clinical monitoring frameworks.

Conclusion

This study introduced a multiclass Convolutional Neural Network (CNN) framework for predicting individual sleep quality based on the integration of physiological and behavioral features collected from wearable and smartphone sensors. The proposed model successfully classified sleep quality into four levels (poor, fair, good, and excellent) with superior accuracy and reliability compared to traditional machine learning and unimodal deep learning baselines.

The multimodal CNN architecture demonstrated that combining physiological signals such as heart rate, HRV, and skin temperature with behavioral patterns including daily activity, sedentary behavior, and smartphone usage provides a richer representation of sleep-related dynamics. The resulting model achieved a macro-F1 score of 0.76 and a macro-AUROC of 0.87, outperforming all benchmark models while maintaining strong calibration (ECE = 0.064). The findings highlight three major contributions:

1. Multiclass prediction capability: Unlike prior binary sleep-quality models, this work introduces a four-level classification that better reflects real-world sleep variability.
2. Multimodal fusion effectiveness: The integration of physiological and behavioral signals enhances model robustness and interpretability, offering a holistic view of sleep health.
3. Interpretable deep learning: Grad-CAM visualizations reveal physiologically meaningful features such as stable HRV patterns and consistent bedtime routines that align with established sleep science.

Beyond methodological advances, this research provides a foundation for personalized sleep monitoring systems capable of delivering real-time feedback and behavioral recommendations. The results indicate the potential for integrating such models into digital health platforms, enabling proactive management of sleep hygiene and related lifestyle interventions.

However, several limitations remain. Sleep quality labels derived from the PSQI are subjective and may not perfectly align with physiological ground truth. Moreover, the dataset was collected from a limited range of wearable devices, which may restrict model generalization across populations and hardware ecosystems. Future studies should therefore incorporate objective polysomnography data, cross-device adaptation, and longitudinal learning frameworks to enhance scalability and clinical applicability.

In summary, the proposed CNN multikategori framework represents a significant step toward data-driven, interpretable, and personalized assessment of sleep quality. It bridges the gap between physiological measurement and behavioral context, supporting a new generation of intelligent health-monitoring technologies focused on sleep wellness and preventive medicine.

References

[1] M. Kline, "Sleep quality and health outcomes," *Sleep Medicine Reviews*, vol. 45, pp. 1–10, 2019.

[2] D. Buysse, “Sleep health: Can we define it? Does it matter?,” *Sleep*, vol. 37, no. 1, pp. 9–17, 2014.
[3] J. F. Reeder and T. L. Brown, “Wearable sensors and the future of sleep research,” *Frontiers in Neuroscience*, vol. 15, 2021.
[4] Gupta et al., “Machine learning approaches for sleep quality prediction: A review,” *IEEE Access*, vol. 10, pp. 11230–11245, 2022.
[5] Y. LeCun et al., “Deep learning,” *Nature*, vol. 521, pp. 436–444, 2015.
[6] X. Zhang, et al., “Sleep stage classification using CNN on EEG signals,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 6, pp. 1421–1431, 2020.
[7] S. Li, et al., “Behavioral feature learning for health monitoring using deep CNNs,” *Sensors*, vol. 23, no. 3, 2023.
[8] W. Chen, et al., “Multiclass sleep quality assessment using deep convolutional neural networks,” *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 4, pp. 1905–1916, 2023.
[9] D. J. Buysse, “Sleep health: Can we define it? Does it matter?,” *Sleep*, vol. 37, no. 1, pp. 9–17, 2014, doi: 10.5665/sleep.3298.
[10] Sathyanarayana, S. Joty, L. Fernandez-Luque, F. Ofli, J. Srivastava, A. Elmagarmid, T. Arora, and S. Taheri, “Sleep Quality Prediction From Wearable Data Using Deep Learning,” *JMIR mHealth and uHealth*, vol. 4, no. 4, e125, 2016, doi: 10.2196/mhealth.6562.
[11] Supratak, H. Dong, C. Wu, and Y. Guo, “DeepSleepNet: A Model for Automatic Sleep Stage Scoring Based on Raw Single-Channel EEG,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1998–2008, Nov. 2017, doi: 10.1109/TNSRE.2017.2721116.
[12] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, May 2015, doi: 10.1038/nature14539.
[13] H. Korkalainen *et al.*, “Deep learning enables sleep staging from photoplethysmogram for patients with suspected sleep apnea,” *Sleep*, vol. 43, no. 11, zssa098, 2020, doi: 10.1093/sleep/zsaa098.
[14] M. R. Patterson, A. A. S. Nunes, D. Gerstel, R. Pilkar, T. Guthrie, A. Neishabouri, and C. C. Guo, “40 years of actigraphy in sleep medicine and current state of the art algorithms,” *NPJ Digital Medicine*, vol. 6, no. 51, 2023, doi: 10.1038/s41746-023-00802-1.
[15] Di Credico *et al.*, “Predicting Sleep Quality through Biofeedback: A Machine Learning Approach Using Heart Rate Variability and Skin Temperature,” *Clocks & Sleep*, vol. 6, no. 3, pp. 322–337, 2024, doi: 10.3390/clockssleep6030023.
[16] Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, “On Calibration of Modern Neural Networks,” in *Proc. 34th Int. Conf. Machine Learning (ICML)*, PMLR vol. 70, pp. 1321–1330, 2017.
[17] J. K. Phillips *et al.*, “Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing,” *Scientific Reports*, vol. 7, no. 3216, 2017, doi: 10.1038/s41598-017-03171-4.
[18] L. Zhuang, R. Zhang, and L. Chen, “Intelligent automatic sleep staging model based on CNN for medical diagnosis,” *Frontiers in Public Health*, vol. 10, 946833, 2022, doi: 10.3389/fpubh.2022.946833.
[19] H. Yue, X. Yang, and Z. Peng, “Research and application of deep learning-based sleep staging: current status and prospects,” *Sleep Medicine Reviews*, 2024, doi: 10.1016/j.smrv.2024.101897.
[20] Y. Tai *et al.*, “Association between circadian skin temperature rhythms and sleep quality,” *Sleep*, vol. 46, no. 8, 2023, doi: 10.1093/sleep/zsad001.

[21] X. Jiang, C. Guo, H. Chen, and J. Luo, “Convolutional neural network based on photoplethysmography for sleep analysis,” *Frontiers in Neuroscience*, vol. 17, 1222715, 2023, doi: 10.3389/fnins.2023.1222715.
[22] Zhong *et al*., “Electronic Screen Use and Sleep Duration and Timing in Adolescents,” *JAMA Network Open*, vol. 8, no. 7, e****, 2025. (Associations of pre-bed screen use with poor sleep quality.)
[23] S. Schrempft, K. Bayes, and A. Steptoe, “Associations between bedtime media use and sleep outcomes: a population-based cohort,” *Sleep Medicine*, 2024, doi: 10.1016/j.sleep.2024.08.012.