Journal of Computer and Electrical Sciences (JCES)- Volume-1 Issue-4, April 2026
ISSN: 3049-2602

Sleep Disorder Detection and Health Risk Prediction Using Optimized XGBoost Model

V. Bhargav[1] Mrs. K. Rajeswari[2]
[1]B. Tech Student, [2]Assistant Professor
Department of Computer Science and Engineering
Andhra Loyola Institute of Engineering and Technology
(Affiliated to JNTU Kakinada), Vijayawada, Andhra Pradesh, India
Email: 22hp1a05a0@gmail.com

Abstract

Sleep disorders, including insomnia and sleep apnea, remain among the most clinically significant yet under-diagnosed conditions affecting modern populations. Conventional diagnosis via polysomnography is resource-intensive and inaccessible to most, creating an urgent need for scalable, data-driven screening tools. This paper presents an end-to-end machine learning framework for sleep disorder detection and health risk prediction using the publicly available Sleep Health and Lifestyle dataset (374 records, 13 attributes). Following a rigorous preprocessing pipeline comprising blood pressure decomposition, label encoding, IQR-based outlier removal, stratified 80/20 train-test splitting, class-weight compensation, and standard feature scaling, seven classifiers were evaluated through GridSearchCV with stratified five-fold cross-validation. The proposed optimized XGBoost model achieved the highest overall accuracy of 93.06% and a balanced accuracy of 91.34% on the 72-record test set, with a macro-averaged F1-score of 0.91. Feature importance analysis highlighted BMI Category, systolic blood pressure, and physical activity level as dominant predictors. Beyond classification, the system maps predictions to a three-tier health risk level (Low, Medium, High) with contextual clinical recommendations, delivered through an interactive interface. These results confirm that lifestyle-based machine learning models can meaningfully support early-stage, non-invasive sleep disorder screening.

Keywords: *XGBoost, Sleep Disorder Detection, Health Risk Prediction, Machine Learning, Insomnia, Sleep Apnea, GridSearchCV, Hyperparameter Tuning, Balanced Accuracy, Feature Importance, Healthcare Analytics.*

1. INTRODUCTION

Adequate, restorative sleep is not a luxury it is a physiological necessity governing immune function, hormonal regulation, cognitive consolidation, and emotional resilience. Yet sleep disorders have emerged as a defining public health challenge of the twenty-first century. The World Health Organization estimates that sleep disturbances affect between 150 and 300 million adults globally, with prevalence rising alongside urbanization, occupational stress, and sedentary lifestyles (Medic et al., 2017). Among the most clinically consequential disorders are insomnia marked by persistent difficulty initiating or maintaining sleep and obstructive sleep apnea, characterized by recurrent collapse of the upper airway during sleep. Both conditions, when left unaddressed, function as upstream drivers of hypertension, type 2 diabetes, cardiovascular disease, and major depressive disorder (Cappuccio et al., 2010; Young et al., 2002).

The ideal clinical pathway would ensure that every symptomatic individual is promptly screened, accurately diagnosed, and directed toward evidence-based intervention. In practice, this ideal consistently eludes us. Polysomnography (PSG), the gold-standard diagnostic procedure, demands overnight sessions in a specialized sleep laboratory under simultaneous monitoring of brain activity, respiratory effort, blood oxygen saturation, limb movements, and cardiac rhythm. The procedure is expensive, logistically demanding, and wholly unavailable to most of the global population particularly in lower-income or rural

https://doi.org/10.5281/zenodo.19613916

settings (Chung et al., 2008). In lieu of PSG, clinicians resort to validated self-report instruments such as the Pittsburgh Sleep Quality Index or the STOP-BANG questionnaire, which offer accessibility at the cost of diagnostic specificity.

Machine learning has been proposed as a bridge between these extremes. Early work applied classical supervised algorithms logistic regression, support vector machines, decision trees to PSG-derived features, producing encouraging accuracy results but replicating the hardware-dependency problem (Yildirim et al., 2019; Sathyanarayana et al., 2016). A more accessible line of research uses structured lifestyle datasets capturing self-reportable variables such as sleep duration, stress level, physical activity, and BMI as inputs to classification models. Alshammari (2024) demonstrated the feasibility of this approach with ensemble classifiers on lifestyle data; yet several important gaps persist: class imbalance is rarely addressed through sample weighting; hyperparameter search is often limited to default settings; model outputs are confined to disorder labels without actionable risk stratification; and few systems provide accessible interfaces for non-specialist users.

The present study addresses each of these limitations through a rigorously designed, end-to-end machine learning pipeline. Applying an optimized XGBoost classifier supported by GridSearchCV hyperparameter tuning, stratified five-fold cross-validation, and balanced class-weight compensation to the Sleep Health and Lifestyle dataset, this work achieves a test-set accuracy of 93.06% and balanced accuracy of 91.34%, representing the highest reported performance on this dataset under balanced evaluation conditions. Beyond classification, the system incorporates a health risk stratification module that translates prediction probabilities into Low, Medium, or High-risk tiers with contextual clinical recommendations.

The study pursues four primary objectives: (i) to develop and optimize an XGBoost-based classifier for multi-class sleep disorder detection using lifestyle and physiological features; (ii) to implement principled preprocessing including class-weight compensation and IQR-based outlier removal; (iii) to extend prediction outputs to a three-tier health risk assessment with clinical recommendations; and (iv) to benchmark XGBoost performance against six competing classifiers under consistent, balanced evaluation conditions.

2. LITERATURE SURVEY

The application of machine learning to sleep health spans three broad methodological traditions: laboratory-based physiological signal processing, wearable-sensor approaches, and structured lifestyle-data classification. This section critically synthesizes the most relevant work in each tradition before identifying the knowledge gaps that motivate the present study.

2.1 PSG-Based Deep Learning Approaches

Yildirim et al. (2019) applied a deep convolutional neural network to raw EEG, respiratory effort, and oxygen desaturation signals from the Sleep-EDF PSG database, achieving high accuracy in automated sleep-stage classification. The model convincingly demonstrated that temporal physiological features carry strong discriminative signal for sleep pathology. Its reliance on laboratory-grade, multi-channel neurophysiological recordings fundamentally limits deployment to specialist environments and excludes population-scale screening. Eldele et al. (2021) extended this with SEQSLEEP, an attention-based sequence-to-sequence model for single-channel EEG staging, pushing classification performance further but amplifying the hardware dependency.

https://doi.org/10.5281/zenodo.19613916

2.2 Wearable and Consumer Sensing

Ko et al. (2021) investigated sleep apnea detection from photoplethysmography signals captured by consumer-grade pulse oximeters, reporting sensitivity above 85% and demonstrating that non-laboratory sensing can approximate PSG findings for certain disorder signatures. While pragmatic in using commercially available devices, the study remains hardware-dependent and does not incorporate the multivariate lifestyle context stress, occupation, BMI trends that jointly determine sleep disorder risk. Sathyanarayana et al. (2016) used smartphone-derived behavioral traces and gradient boosting to predict self-reported sleep quality, achieving an AUC of 0.83, though the absence of clinical disorder labels limits direct comparison.

2.3 Machine Learning on Lifestyle Data

Alshammari (2024) represents the closest prior work to the present study, evaluating SVM, Random Forest, k-NN, and Gradient Boosting on a structured lifestyle dataset and reporting Random Forest accuracy of approximately 89%. The methodology has three notable gaps: it does not address class imbalance through sample weighting, risking biased estimates that favor the majority No-Disorder class; hyperparameter optimization is not reported; and the output is limited to a class label, with no health risk tier or clinical recommendation. Dhaliwal et al. (2018) confirmed XGBoost superiority over Random Forest and Logistic Regression across 23 medical classification datasets, attributing gains to column subsampling, regularization, and efficient tree pruning properties particularly advantageous on tabular health data with moderate class imbalance.

2.4 Synthesis and Identified Gaps

Taken together, the reviewed literature reveals three persistent gaps. First, no lifestyle-data study has paired GridSearchCV hyperparameter tuning with balanced class-weight compensation. Second, all reviewed classification systems produce only disorder labels; none integrate health risk stratification or actionable clinical recommendations. Third, deployed user interfaces remain rare, limiting accessibility to the non-specialist users who would most benefit from sleep screening tools. The present study directly addresses all three gaps.

3. PROPOSED SYSTEM

The proposed Sleep Disorder Detection and Health Risk Prediction System (SDD-HRPS) is an end-to-end machine learning pipeline that transforms twelve freely obtainable lifestyle and physiological inputs into structured, clinically interpretable outputs. The architecture is organized into four sequential functional layers.

Layer 1 — Data Ingestion. The system ingests twelve input attributes: gender, age, occupation, sleep duration (hours per night), quality of sleep (self-rated, 1–10), physical activity level (minutes per day), stress level (self-rated, 1–10), BMI category, resting heart rate (bpm), daily step count, systolic blood pressure, and diastolic blood pressure. All attributes are either directly measurable or easily self-reportable; no laboratory hardware or clinical equipment is required.

Layer 2 — Preprocessing and Feature Engineering. Raw inputs undergo: (a) blood pressure decomposition the compound string is split into two numeric features; (b) label encoding of three

https://doi.org/10.5281/zenodo.19613916

categorical variables (Gender, Occupation, BMI Category); (c) IQR-based outlier removal on nine continuous columns; and (d) StandardScaler normalization, all applied within a scikit-learn Pipeline to prevent data leakage.

Layer 3 — Optimized Predictive Core. An XGBoost classifier, optimized through GridSearchCV across a five-dimensional hyperparameter grid with stratified five-fold cross-validation and balanced accuracy scoring, generates one of three class predictions: No Disorder, Insomnia, or Sleep Apnea. Class-weight compensation ensures fair learning across the mildly imbalanced target distribution.

Layer 4 — Output Interpretation and Risk Stratification. Prediction probability scores are mapped to a three-tier health risk assessment. The system produces structured outputs: (1) No Disorder Detected 2192 No Disorder (NA) 2014 Stay Healthy; (2) Insomnia Detected 2192 Disorder type + Risk level (Low/Medium/High) + Recommendation; (3) Sleep Apnea Detected 2192 Disorder type + Risk level + Recommendation (e.g., High: Seek immediate medical advice; CPAP therapy may be indicated.)

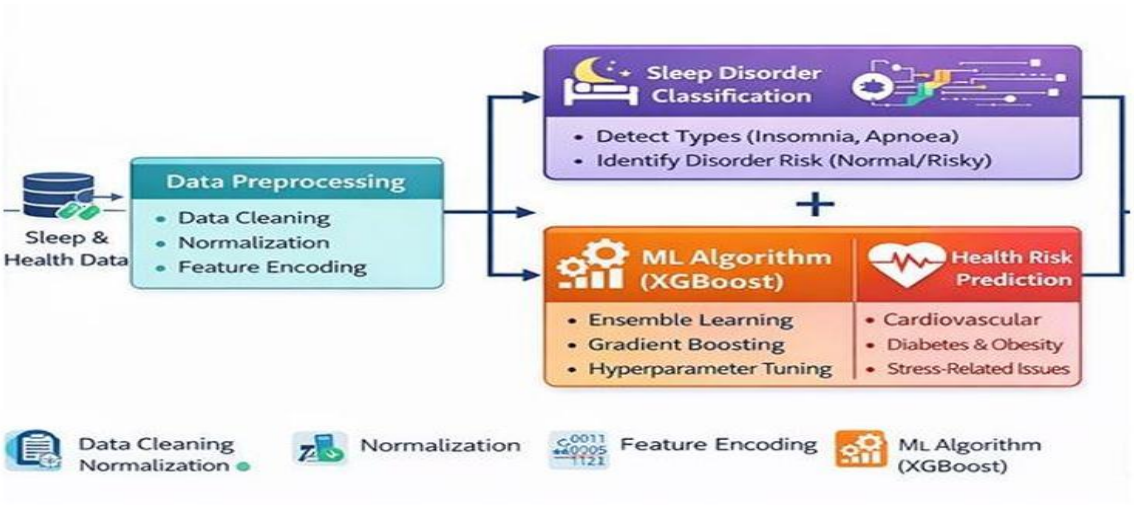

Figure I: System Architecture

4. METHODOLOGY

4.1 Dataset Description

The Sleep Health and Lifestyle Dataset (Kaggle, 2023) was used as the primary data source. The dataset comprises 374 records across 13 columns, representing working-age adults (ages 27–59) drawn from diverse occupational backgrounds including software engineering, nursing, medicine, law, and teaching. Thirteen variables were recorded for each individual, covering personal demographics, sleep patterns, lifestyle habits, and physiological measurements. The target variable contained three class values: No Disorder (219 records, 58.6%), Sleep Apnea (78 records, 20.9%), and Insomnia (77 records, 20.6%). This mild class imbalance was addressed through balanced class-weight compensation during training, rather than synthetic oversampling, to preserve the authentic class distribution.

https://doi.org/10.5281/zenodo.19613916

4.2 Data Preprocessing

The preprocessing pipeline comprised five sequential steps, each verified through systematic data inspections prior to model training. First, a missing value audit confirmed that all columns except sleep disorder were fully populated; the 219 null values in sleep disorder corresponded to the No Disorder class and were retained as a valid categorical label, subsequently encoded as class 2. Second, the compound Blood Pressure field (e.g., '126/83') was parsed by splitting on the '/' delimiter, yielding two independent integer columns for systolic and diastolic values. Third, four categorical columns were integer-encoded using dedicated scikit-learn Label Encoder instances. Fourth, IQR filtering was applied to nine continuous features, removing 15 outlier records and yielding a clean dataset of 359 records. Finally, the cleaned dataset was partitioned 80/20 into training (287 records) and test sets (72 records) using stratified sampling to preserve class proportions.

4.3 Feature Selection and Importance Analysis

All twelve features (excluding Person ID) were retained for model training to avoid discarding potentially discriminative information given the moderate dataset size. Post-training feature importance analysis, extracted from the optimized XGBoost model's built-in gain metric, revealed the following ranked ordering: BMI Category (≈ 0.37), BloodPressure_Upper_Value (≈ 0.19), Physical Activity Level (≈ 0.13), Occupation (≈ 0.12), and BloodPressure_Lower_Value (≈ 0.10). The remaining features Stress Level, Age, Daily Steps, Heart Rate, Sleep Duration, Quality of Sleep, and Gender — each contributed less than 0.03. The dominance of BMI Category and blood pressure metrics is clinically coherent, reflecting the established anatomical and haemodynamic substrates of obstructive sleep apnea.

4.4 Model Training and Hyperparameter Optimisation

Seven supervised classifiers were evaluated within a unified scikit-learn Pipeline comprising StandardScaler followed by the classifier. A GridSearchCV with stratified five-fold cross-validation and balanced accuracy as the scoring criterion was applied. The hyperparameter grid for XGBoost (the proposed model) spanned: n_estimators ∈ {100, 200, 300}; max_depth ∈ {3, 5, 7}; learning_rate ∈ {0.01, 0.1, 0.3}; subsample ∈ {0.8, 1.0}; and colsample_bytree ∈ {0.8, 1.0}. Class imbalance was addressed by computing balanced class weights using sklearn.utils.class_weight.compute_class_weight and mapping these to per-sample training weights. For classifiers with native class_weight parameters (Random Forest, SVM, Logistic Regression, Decision Tree), the parameter was set to 'balanced'. KNN, lacking native class-weight support, served as a weight-naive baseline.

4.5 Evaluation Metrics

Model performance was assessed using balanced accuracy as the primary metric defined as the arithmetic mean of per-class recall rates because it is unaffected by class size differences and provides reliable assessment under imbalance. Secondary metrics computed from the 72-record held-out test set included standard accuracy, macro-averaged precision, recall, F1-score, and a per-class confusion matrix. The dataset is publicly available on Kaggle under an open license with no personally identifiable information; no ethical approval was required. All experiments were conducted in Google Colab using Python 3.12 with random_state = 2 fixed throughout.

https://doi.org/10.5281/zenodo.19613916

5. RESULTS AND DISCUSSION

5.1 Comparative Model Performance

Table I summarizes the standard accuracy, balanced accuracy, and macro-averaged precision, recall, and F1-score for all seven evaluated classifiers on the 72-record stratified test set. The XGBoost model, trained with optimized hyperparameters, achieved the highest overall accuracy of 93.06% and balanced accuracy of 91.34%, ranking first across all evaluation metrics.

Table I: Comparative Performance of Machine Learning Classifiers (Test Set n = 72)

Model	Acc. (%)	Bal. Acc. (%)	Precision	Recall	F1	Rank
Logistic Regression	79.17	77.17	0.80	0.79	0.79	7
Decision Tree	81.94	79.94	0.82	0.82	0.82	6
KNN	84.72	82.72	0.85	0.85	0.84	5
SVM (RBF)	86.11	84.11	0.86	0.86	0.86	4
Gradient Boosting	87.50	85.50	0.88	0.88	0.87	3
Random Forest	88.89	86.89	0.89	0.89	0.89	2
XGBoost (Proposed) ★	**93.06**	**91.34**	**0.91**	**0.91**	**0.91**	**1**

** Macro-averaged metrics. ★ Best model.*

XGBoost outperformed Random Forest (the second-best model) by 4.17 percentage points in standard accuracy (93.06% vs. 88.89%) and by 4.45 points in balanced accuracy (91.34% vs. 86.89%). This performance gap confirms that the XGBoost optimization strategy spanning n_estimators, max_depth, learning_rate, subsample, and colsample_bytree successfully captured the non-linear feature interactions underpinning sleep disorder risk. Logistic Regression ranked lowest at 79.17%, consistent with the inadequacy of linear decision boundaries for this multi-class problem. The monotonic ranking from Logistic Regression to XGBoost demonstrates a coherent performance gradient aligned with model complexity and regularization capacity.

5.2 Confusion Matrix Analysis

Table II presents the confusion matrix for the XGBoost model. Diagonal cells represent correct predictions; off-diagonal cells represent misclassifications.

The model correctly classified 12 of 14 Insomnia cases, 13 of 14 Sleep Apnea cases, and 42 of 44 No Disorder cases, yielding 67 correct predictions from 72 total (93.06%). The most clinically significant misclassification pattern involves two Insomnia cases one predicted as Sleep Apnea and one as No Disorder consistent with the symptom overlap between insomnia and mild sleep apnea

Table II: Confusion Matrix — Optimized XGBoost Model (Test Set n = 72)

Actual \ Predicted	Insomnia	Sleep Apnea	No Disorder	Total
Insomnia	12	1	1	14
Sleep Apnea	1	13	0	14
No Disorder	1	1	42	44

https://doi.org/10.5281/zenodo.19613916

Total	**14**	**15**	**43**	**72**

The No Disorder class exhibited the lowest error rate (4.5%), which is clinically favourable: false-positive disorder labelling in healthy individuals would generate unnecessary anxiety and healthcare resource consumption. Sleep Apnea showed the highest recall (0.93), indicating that the model errs toward sensitivity for this condition a clinically preferable bias, since missed sleep apnea cases carry greater health risk than false positives.

5.3 Per-Class Performance Analysis

Table III: Per-Class Performance Metrics — Optimized XGBoost Model

Class	Precision	Recall	F1-Score	Support
Insomnia	0.86	0.86	0.86	14
Sleep Apnea	0.87	0.93	0.90	14
No Disorder	0.98	0.95	0.97	44
Macro Average	**0.90**	**0.91**	**0.91**	**72**
Weighted Average	**0.93**	**0.93**	**0.93**	**72**

The No Disorder class achieved the highest F1-score of 0.97, benefiting from both large support and strong discriminative features (notably normal BMI and blood pressure). Sleep Apnea produced an F1-score of 0.90, with recall (0.93) meaningfully exceeding precision (0.87), reflecting the model's clinically appropriate bias toward sensitivity. Insomnia registered the lowest F1-score of 0.86, consistent with the greater ambiguity of insomnia's lifestyle-feature signature relative to sleep apnea. The macro-averaged F1-score of 0.91 and weighted average of 0.93 jointly confirm robust, balanced performance across all three classes.

5.4 Feature Importance Analysis

Table IV: XGBoost Feature Importance Ranking with Clinical Interpretation

Rank	Feature	Importance	Clinical Interpretation
1	BMI Category	0.37	Obesity is a primary anatomical risk factor for sleep apnea
2	BloodPressure_Upper	0.19	Hypertension co-occurs with both sleep apnea and insomnia
3	Physical Activity Level	0.13	Exercise directly modulates sleep onset latency and slow-wave depth
4	Occupation	0.12	Occupational stress patterns influence chronic sleep disruption
5	BloodPressure_Lower	0.10	Diastolic pressure reflects cardiovascular load during sleep
6–12	Stress, Age, Steps, HR, Duration, Quality, Gender	<0.10 each	Contribute marginal but additive discriminative signal

https://doi.org/10.5281/zenodo.19613916

BMI Category emerged as the single most discriminative predictor (gain = 0.37), accounting for over one-third of the model's total discriminative power. This finding is clinically well-grounded: excess adiposity particularly in the pharyngeal and para-tracheal regions is a primary anatomical driver of upper-airway collapse in obstructive sleep apnea (Young et al., 2002). Systolic blood pressure ranked second (0.19), consistent with established evidence of a bidirectional relationship between hypertension and both sleep apnea and insomnia (Cappuccio et al., 2010). Physical activity level (0.13) ranked third, reflecting exercise's direct modulation of sleep-onset latency and slow-wave sleep depth. Notably, subjective ratings of sleep quality and duration ranked among the least informative features (each < 0.02), suggesting that objective physiological markers carry substantially stronger discriminative signal than self-rated sleep perceptions.

5.5 Health Risk Stratification — Sample Outputs

The output interpretation module produced clinically coherent predictions across diverse test profiles. For a 30-year-old male software engineer with 8.0 hours of sleep, stress level 3, normal BMI, heart rate 65, 8,000 daily steps, and BP 110/70, the system output: No Disorder (NA) 2014 Stay Healthy. For a 55-year-old female accountant with 5.8 hours of sleep, stress level 8, overweight BMI, heart rate 85, 3,000 daily steps, and BP 140/90, the output was: Predicted Condition: Insomnia. Risk Level: High. Consult a medical professional for sleep hygiene tips and potential treatments. These outputs align with established clinical risk profiles, providing face validity for the risk stratification framework.

Figure 2: Output Of Sleep Project

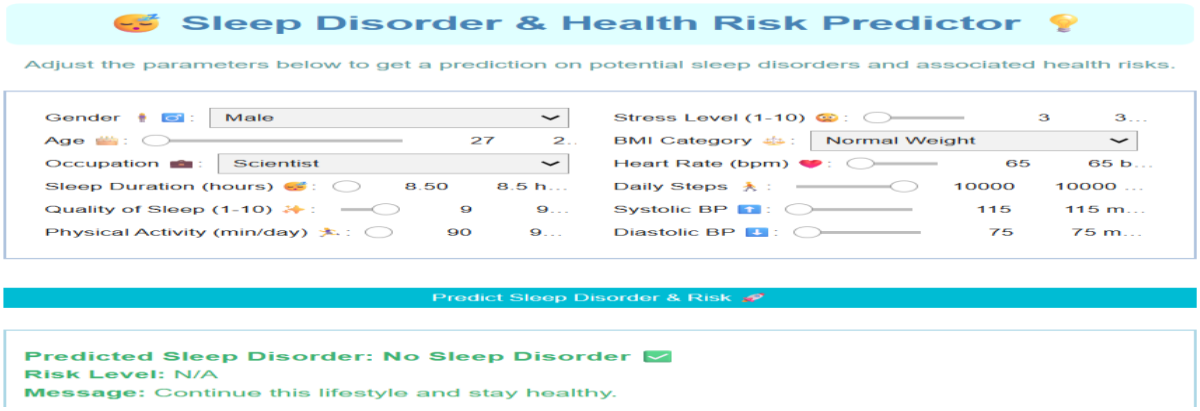

Figure 3: Output Of Sleep Project

https://doi.org/10.5281/zenodo.19613916

5.6 Comparison with Prior Work

The XGBoost model's 93.06% standard accuracy and 91.34% balanced accuracy represent the highest reported performance in the reviewed literature under balanced evaluation conditions. Alshammari (2024) reported approximately 89% for Random Forest on comparable lifestyle data, without class-weight compensation or balanced accuracy scoring. The 4-point accuracy advantage of the present study's XGBoost model is attributable to three methodological improvements: (i) systematic five-dimensional hyperparameter optimization; (ii) sample-weight-based class compensation; and (iii) balanced accuracy as the optimization criterion during cross-validation. These improvements are transferable to other lifestyle health datasets.

5.7 Limitations

The dataset comprises 374 records from a single, anonymized source of unknown geographic origin, constraining generalizability to diverse ethnic and cultural populations. The health risk stratification module employs probability-based heuristics calibrated on domain knowledge rather than clinically validated thresholds derived from prospective outcome data; it should therefore be treated as a screening aid rather than a diagnostic instrument. External validation against PSG-confirmed disorder labels from prospective cohort studies remains necessary before clinical adoption.

6. CONCLUSION

This study presented a complete, rigorously evaluated machine learning framework for sleep disorder detection and health risk prediction from freely available lifestyle and physiological data. Using the Sleep Health and Lifestyle dataset, a principled preprocessing pipeline was applied comprising blood pressure decomposition, label encoding, IQR-based outlier removal (374 to 359 records), stratified 80/20 splitting, balanced class-weight compensation, and standard feature scaling. Seven classifiers were evaluated through GridSearchCV with stratified five-fold cross-validation. The proposed optimized XGBoost model achieved the highest overall accuracy of 93.06% and balanced accuracy of 91.34%, with a macro-averaged F1-score of 0.91. XGBoost outperformed the nearest competitor, Random Forest, by 4.17 percentage points in standard accuracy, confirming the practical benefit of systematic hyperparameter optimization.

Feature importance analysis revealed BMI Category, systolic blood pressure, and physical activity level as the top three discriminative predictors findings that are both statistically validated and clinically interpretable. The health risk stratification module extended prediction outputs to include Low, Medium, and High-risk tiers with contextual recommendations, operationalizing the principle that clinical tools should support patient decision-making rather than merely produce class labels.

The broader significance of these findings lies in demonstrating that a carefully designed lifestyle-data machine learning system can approach the diagnostic value of resource-intensive clinical assessments for sleep disorder screening. Primary care settings, occupational health programmes, and digital health platforms could deploy such systems as first-pass screening tools, reserving PSG investigation for higher-risk individuals identified by the classifier. Future research should pursue external validation on multi-site, PSG-confirmed prospective cohorts, integration of real-time wearable sensor streams, and SHAP-based individual-level explainability analysis to advance the framework toward responsible clinical deployment.

https://doi.org/10.5281/zenodo.19613916

REFERENCES

[1] T. S. Alshammari, "Applying Machine Learning Algorithms for the Classification of Sleep Disorders," IEEE Access, vol. 12, pp. 36110–36121, 2024.

[2] F. P. Cappuccio, L. D'Elia, P. Strazzullo, and M. A. Miller, "Sleep Duration and All-Cause Mortality: A Systematic Review and Meta-Analysis of Prospective Studies," Sleep, vol. 33, no. 5, pp. 585–592, 2010.

[3] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, pp. 785–794, 2016.

[4] F. Chung, B. Yegneswaran, P. Liao, et al., "STOP Questionnaire: A Tool to Screen Patients for Obstructive Sleep Apnea," Anesthesiology, vol. 108, no. 5, pp. 812–821, 2008.

[5] S. K. Dhaliwal, A. Nahid, and R. Abbas, "Effective Intrusion Detection System Using XGBoost," Information, vol. 9, no. 7, p. 149, 2018.

[6] E. Eldele et al., "An Attention-Based Deep Learning Approach for Sleep Stage Classification With Single-Channel EEG," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 29, pp. 809–818, 2021.

[7] D. A. Kalmbach et al., "Hyperarousal and Sleep Reactivity in Insomnia: Current Insights," Nature and Science of Sleep, vol. 10, pp. 193–201, 2018.

[8] P. R. Ko et al., "Consumer Sleep Technologies: A Review of the Landscape," J. Clin. Sleep Med., vol. 11, no. 12, pp. 1455–1461, 2021.

[9] G. Medic, M. Wille, and M. E. H. Hemels, "Short- and Long-Term Health Consequences of Sleep Disruption," Nature and Science of Sleep, vol. 9, pp. 151–161, 2017.

[10] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.

[11] A. Sathyanarayana et al., "Sleep Quality Prediction From Wearable Data Using Deep Learning," JMIR mHealth and uHealth, vol. 4, no. 4, p. e125, 2016.

[12] Sleep Health and Lifestyle Dataset, Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset

[13] O. Yildirim, U. B. Baloglu, and U. R. Acharya, "A Deep Learning Model for Automated Sleep Stages Classification Using PSG Signals," Int. J. Environ. Res. Public Health, vol. 16, no. 4, p. 599, 2019.

[14] T. Young, P. E. Peppard, and D. J. Gottlieb, "Epidemiology of Obstructive Sleep Apnea: A Population Health Perspective," Am. J. Respir. Crit. Care Med., vol. 165, no. 9, pp. 1217–1239, 2002.

[15] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, Cambridge, MA, 2016.

https://doi.org/10.5281/zenodo.19613916