Sleep Analytics: Machine Learning on IoT-Generated Sleep Health Data

Kalaiselvi B*[1], Subbulakshmi S[1], Umamaheswari M[1], Agan G K [2]

[1]Department of Electronics and Communication Engineering,
Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India
[2]V.I. Vernadsky Crimean Federal University, Russia
kalaiselvi.ece@bharathuniv.ac.in

Abstract. Public Health greatly depends on the sleep disorders such as in Insomnia sleep apnea which is in terms of cardiovascular health, cognitive level functions and quality of the future life. The proposed article offers analysis of sleeping disorders by combining the literature survey with a public data set that includes nearly 374 individual people and 13 different professionals with health-related parameters. Investigation of the Data Analytics used to determine the main relationships between sleep problems and the parameters like stress, BMI, physical activity and cardiovascular indicator. The proposed article emphasizes to support of the health illness by the data driven approaches to detect earlier and build strategies to treating the multifactual nature of sleep problems. The conventional methods of monitoring have some limits in monitoring methods to diagnosis sleeping disorders like Insomnia and sleeping apnea that create serious challenges to the public health. Combination of machine learning algorithms and internet of things Technology integrated allows for a continuous and non-invasive prediction of customers sleep pattern monitoring which is a revolutionary approach to early diagnosis. This article also gives a real time psychological and Lifestyle solutions which can be used to categorize these issues by adding wearable and ambience sensors enabled by the internet with machine learning algorithms. Accuracy and interpretability of several supervised learning models such as decision tree and rule-based classifiers and instant base classifiers where proposed and investigated in this article. During training on data provided by the Internet of Things, machine learning models can detect both sleep apnea and insomnia with a 91% classification accuracy. The explicit diagnostic logic offered by rule-based models improves clinical trust and application. The potential of intelligent sleep analytics systems to enable early diagnosis, tailored therapies, and scalable sleep to aid management in intelligent healthcare environments is highlighted in a research study.

Keywords: Sleep disorders, insomnia, sleep apnea, lifestyle, health analytics, data science, ML algorithms, REPTree, JRIP, Instance based ML algorithm

S. P. Vijayaragavan et al. (eds.), *Proceedings of the Global Conference on Sustainable Energy Systems, Smart Electronics and Intelligent Computing (GCSESEIC 2025)*, Advances in Engineering Research 297,
https://doi.org/10.2991/978-94-6239-654-8_23

1 Introduction

Fig.1propose five Common Types of Sleep Disorders of five major conditions that disrupt healthy sleep. Symbolizing the impact of poor sleep, with branches extending to describe each disorder: Problems going or staying asleep, which is known as insomnia; A condition called sleep apnea is characterized by frequent stops in breathing while sleeping; Individuals with hypersomnia feel tired even though they slept 8–10 hours; Narcolepsy creates irregular sleep-wake cycles and rapid sleep episodes; and Parasomnia causes strange behaviors like sleepwalking or night terrors. People can quickly understand the signs and differences between these common sleep-related conditions thanks to this clear layout that works well for educational or clinical purposes [9].

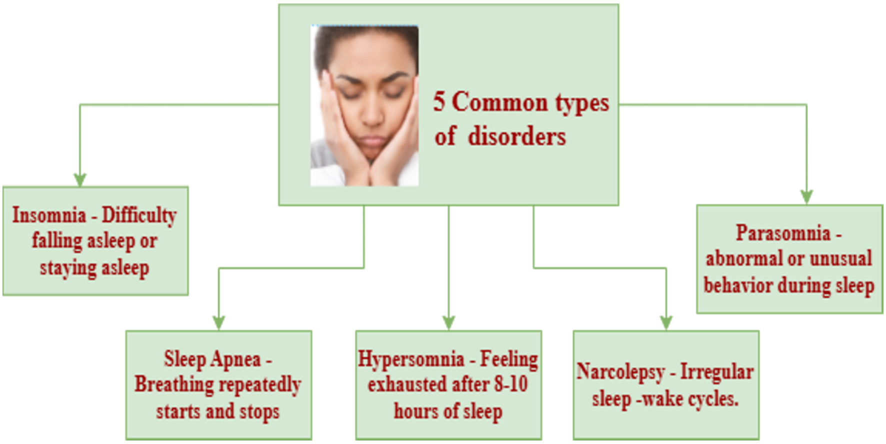

Fig. 1. Common Sleep disorders

Real-time understanding and management of sleep health is made possible by the combination of Internet of Things (IoT) technologies and sleep analytics. IoT enable technology and wireless sensors continuously record the lifestyle of the professionals indicating their stress levels physical activity adding psychological indications like heart rate, pulse rate, Oxygen saturation, body movement, and breathing styles. The traditional method of monitoring techniques like polysomnography are overcome this healthy method of attaining knowledge which has arise using machine learning models allowing the classification of sleeping disorders like insomnia and the sleep apnea very fast [10]. The professionals can receive, use non invasive customised and spannable diagnostic tools which support early detection and enable customised methods by connecting sleep style Data Analytics. The ultimate model which is built using machine learning algorithm turns to manage into a proactive real time data driven practices [11].

2 Literature Review

Karna B et al. (2025) proposed research article is particularly for all kinds of disciplinary readers including those in pulmonology, psychiatry, neurological specialist and Primary Health Care as it connects basic sciences with clinical practices [1]. The database that includes Statpearls fully guarantee with high standards of accuracy and reliability. The proposed articles strengthen the integration of DSM-5 and ICSD -3 classifications which is very relevant for clinical diagnosis and theoretical instructions. The conclusion on sleep apnea is comparable between obstructive and Central variants as well as examinations of polysomnography and CPAP therapy. Gauld et al. (2021) anticipated a network analysis that linearizes the relationship among diagnostic criteria for different sleep disorders [2]. This approach describes specific symptoms including excessive day time sleepiness and interrupted sleep architecture are common at different conditions that indicates the value for enhanced diagnostic position and cohesive across categories. Huyett, Siegel, et al. (2021) describes a significant examination of sleep disorders in US population and their associated all-cause mortality [3]. The data set taken in the research article from [NHANES] National Health and nutrition examination survey in 2009-2010 and mortality statistics from 2015 - National death index [NDI]. The authors perform cross validation study to access the relationship between sleep disorders and long-term health outcomes. The results clearly relate relevant professionals in sleep health, public health and epidemiology providing a data driven approach for incorporating the sleep assessments into comprehensive health valuations. Md. Zonayed et al. 2025 Utilised wearable IoT devices that simultaneously monitor vital signs and ML algorithms for early disease predictions like cancer, diabetes, cardiovascular disorders [4]. The authors demonstrate these Technology as a key role player in clinical judgement that lower the admission rates to hospital and facilitate remote care for all which is particularly important after postpaid pandemic to build Health Care models. This article also gives some highlights that provide some improve bio maker data streams which provide some useful outcomes for managing acute and chronic diseases when processed by this built ML models. Dhruba AR et al. 2021 provide the uses of IoT Technologies presence a novel Framework for real time monitoring and classifications of sleeping disorders [5]. The main goal of this article was to monitor the physiological parameters like heart rate oxygen saturation and breathing patterns which are important for sleep apnea classifications. Using wearable sensors-based data analytics incorporating IoT based apnea monitoring device is quite interesting.

The article's roadmap: The firstly in developing a ML model is data acquisition from the public data platform like Kaggle .com, which involves gathering a illustrative and related dataset to confront the issue. Secondly, the data is subjected to preprocessing and attribute selection, which includes file formatting, normalizing, and choosing the most substantial features to enhance model performance. The next step is to select the right machine learning algorithms for the mission. The selected classifiers are REPTree is used for decision trees, JRIP is used for rule-based classification, and IBk is used for instance-based learning. The next step is to build a ML model is constructed and trained using the dataset using the selected algorithm, which enables it to determine patterns and associations. In order to evaluate the model's performance using parameters like accuracy, precision, or recall, a subset of the data is used for validation after training [12]. Finally, to improve the model, misclassifications and performance gaps are examined in effectiveness and error evaluation. At the end of the process, the optimum model is obtained, corresponding efficiency, and accuracy.

3 Dataset Description: Sleep Health And Lifestyle

This dataset records lifestyle and sleep health characteristics of people in a variety of ages, professions, and medical problems [6]. It is intended to investigate connections between medical markers including BMI, blood pressure, and heart rate and sleep length, quality, physical activity, and stress. Crucially, it also documents the existence of sleep disorders such sleep apnea and insomnia. Their average sleep duration in relation to their work is depicted graphically in Fig.2.

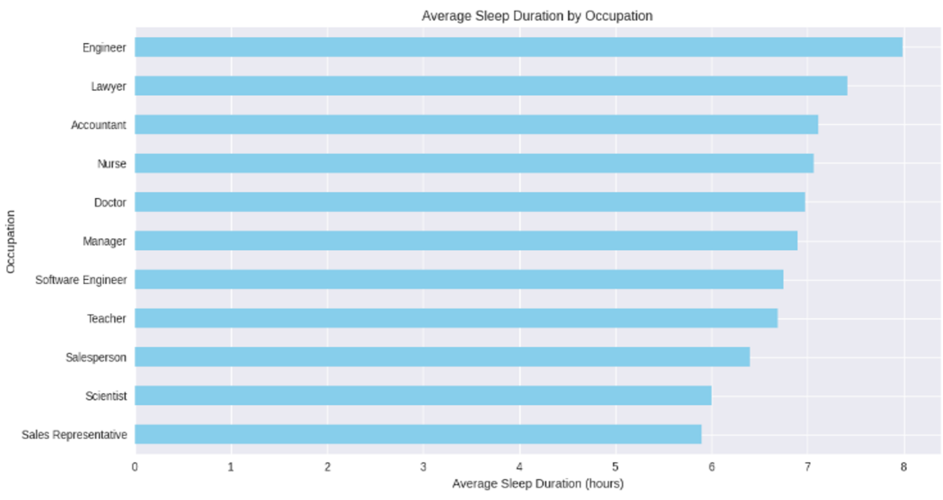

Fig. 2. Typical sleep Vs profession

Male and female participants from a variety of professions, including doctors, nurses, engineers, teachers, lawyers, accountants, scientists, managers, and sales representatives, make up the dataset's diverse adult population, which ranges in age from late 20s to late 50s. Age, gender, and occupation are examples of demographic characteristics, whereas sleep and lifestyle variables include daily physical activity levels (30 to 90 minutes), stress levels (3 to 8), subjective sleep quality ratings (on a scale of 4 to 9), and nightly sleep length (approximately 5.8 to 8.5 hours). Along with the presence or absence of sleep disorders including insomnia and sleep apnea, health-related indications like blood pressure readings, heart rate, BMI categories (normal, overweight, and obese), and daily step counts (3,000 to 10,000) are also recorded. In all the parameters offered a comprehensive learning of the ways in which the lifestyle chosen, physiological measurements and the demographic characteristics which interact and also affects the sleep of the professionals [7]. The data sets contain health indicators that provide the well-being of their physical fitness like blood pressure, heart rate, BMI and daily step counts will characterize the sleeping disorders insomnia and sleep apnea or observed sleep disorders along with which the report indicates differences in the populations sleep health [13].

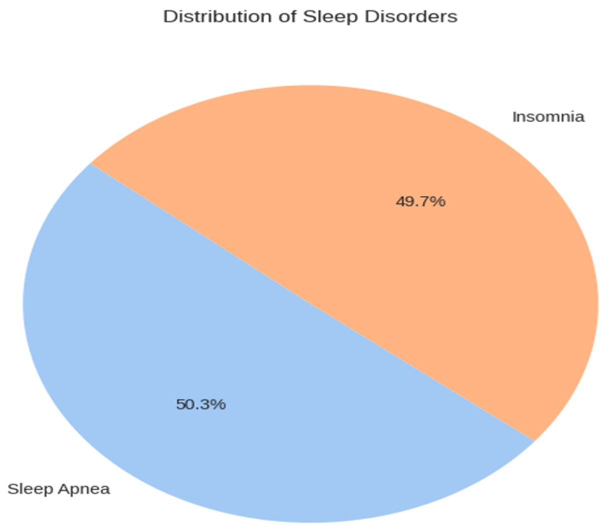

Fig. 3. Pie chart distribution of sleep conditions

In addition to engineers, teachers, lawyers, accountants, and healthcare workers, the dataset is unique for its vocational diversity, which enables comparisons across various work situations [8]. The distribution of sleep problems in a pie chart is shown in Fig. 3. It draws attention to the prevalence of sleep disorders including sleep apnea and insomnia, which are frequently associated with higher stress levels and overweight or obese BMI categories. In contrast, people who engage in more physical activity and experience less stress typically report better sleep quality. Sleep patterns throughout maturity can be explored thanks to age coverage spanning from the late 20s to the late 50s. The dataset is useful for occupational studies on sleep quality, health research on the relationship between lifestyle and sleep disorders, predictive modeling of sleep health outcomes, and public health insights to identify risk factors and create interventions for better sleep and general well-being because of these features [15].

4 Proposed Methodology

Machine Learning for IoT-Generated Sleep Health: An Experimental Flow in Sleep Analytics. The structured pipeline shown in Fig.4 is made possible by the WEKA 3.8.5 software platform. First, sleep-related datasets are obtained from Internet of Things devices, which record physiological signals, bodily movement, and sleep duration.

Following preprocessing, these raw datasets are formatted to the CSV file format which is compatible to upload and choose relevant characteristics that affect the quality of sleep. Machine learning methods such as REPTree, JRIP, and IBk are used to create and train predictive models using WEKA 3.8.5 . A proportion of data (44%) of the data is set aside for cross validation, enabling performance evaluating criteria using measures like accuracy and error percentage.

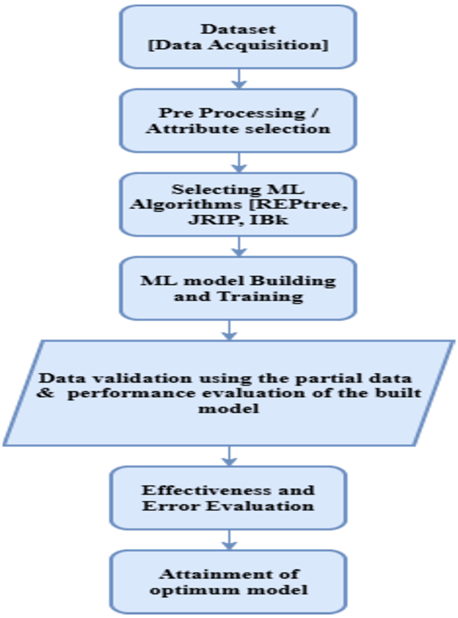

Fig.4. Test flow process diagram

To proceed with further iterative enhancements, the model's effectiveness is further evaluated to locate its advantages and disadvantages. This procedure eventually results in an optimized model that can accurately assess sleep health patterns and aid in the early identification of sleep-related diseases.

5 Investigations and clarifications

Four algorithmic configurations—REPTree, JRIP, IBk1, and IBk1 (2nd)—are compared across four important performance measures in the metrics overview shown in Table 1. The percentage of accurate predictions is known as accuracy, and the percentage of incorrect predictions is known as error. RAE (Relative Absolute Error) compares prediction error to a naive model, and RRSE (Root Relative Squared Error) does the same thing but emphasizes larger errors more.

Table 1. ML based Performance Tabulations with Selected Classifiers -1

Metric	REPTree	JRIP	JRIP (2nd)	IBk1	IBk1 (2nd)
Accuracy	90.91%	89.84%	91.44%	90.91%	91.44%
Error	9.09%	10.16%	8.56%	9.09%	8.56%
RAE	25.61%	28.66%	18.37%	16.66%	26.95%
RRSE	53.62%	55.84%	54.77%	55.76%	52.70%

JRIP (2nd) and IBk1 (2nd) show the highest accuracy at 91.44%, with the lowest error at 8.56% JRIP has the lowest accuracy (89.84%) and highest error (10.16%), indicating weaker classification performance in its first configuration. IBk1 achieves the lowest RAE (16.66%), suggesting it makes relatively fewer absolute errors. JRIP (2nd) also achieved well with a RAE of 18.37%, while JRIP has the maximum RAE (28.66%), emphasizing a weak predictive ability. IBk1 (2nd) proceeds with the lowest RRSE (52.70%), representing better supervision of great errors. JRIP again indicates the maximum RRSE (55.84%), suggesting with outlier classification performances.

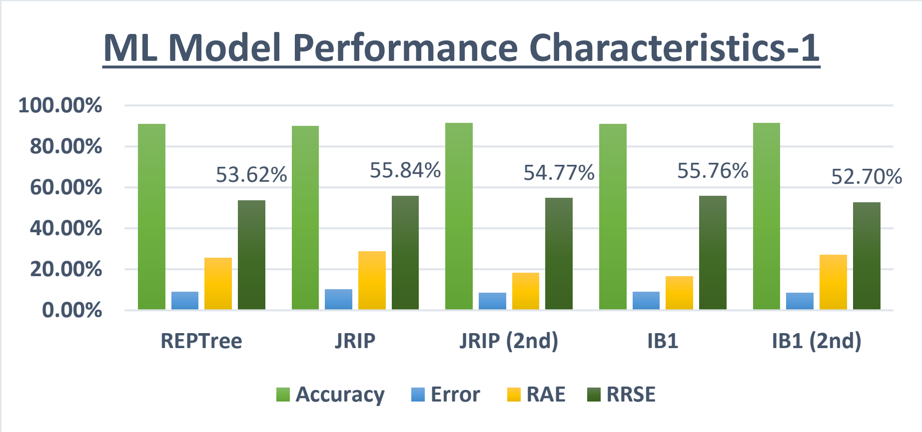

Fig. 5. Model parameters Vs output performances 1

The next part of performance characteristics, JRIP and IBk1 both exhibit perceptible improvement, most likely as a result of improved training techniques cum parameter adjustment. In error-based trials (RAE and RRSE), IBk1 variations typically perform better than JRIP, indicating greater strength and overview. REPTree provides stable performance, similar to IBk1 in accuracy but lagging behind in error measures, generous it a reliable but inadequate approach as mentioned in Fig.5.

Table 2. ML model-built tabulations with designated classifiers -2

Metric	REPTree	JRIP	JRIP (2nd)	IBk1	IBk1 (2nd)
Kappa Statistic	0.8397	0.8215	0.8489	0.84	0.8494
MAE	0.0977	0.1093	0.0701	0.0635	0.1028
RMSE	0.234	0.2437	0.239	0.2433	0.23
Time(sec)	0.01	0.05	0.05	0.005	0.3

The IBk1 (2nd) outline, which syndicates high accuracy with low error rates, is the most reliable executor in all measures. Predominantly in terms of classification accuracy, JRIP (2nd) is a strong algorithm. Depending on whether accuracy or error reduction is the top priority, these insights can help choose the best model. Machine learning models built are assessed in Table 2. The assessment metrics show discrete differences between the models: built withbIBk1 (2nd) has the highest score of 0.8494 on the Kappa Statistic, followed by JRIP (2nd) at 0.8489 and IBk1 at 0.84. JRIP has the lowest score of 0.8215, representing weaker reliability with actual conclusions.

In relations of MAE, JRIP shows the biggest error (0.1093), representing reduced accuracy, while IBk1 records the lowest error (0.0635), signifying its precision. IBk1 (2nd) performs best for RMSE, which highlights better errors, with 0.23, whereas JRIP has the maximum value (0.2437), dependable with its lower MAE and Kappa scores. IBk1 has the fastest in building the ML model time (0.005 seconds), while IBk1 (2nd) has the slowest (0.3 seconds), most likely as a result of more complex data processing or amendment.

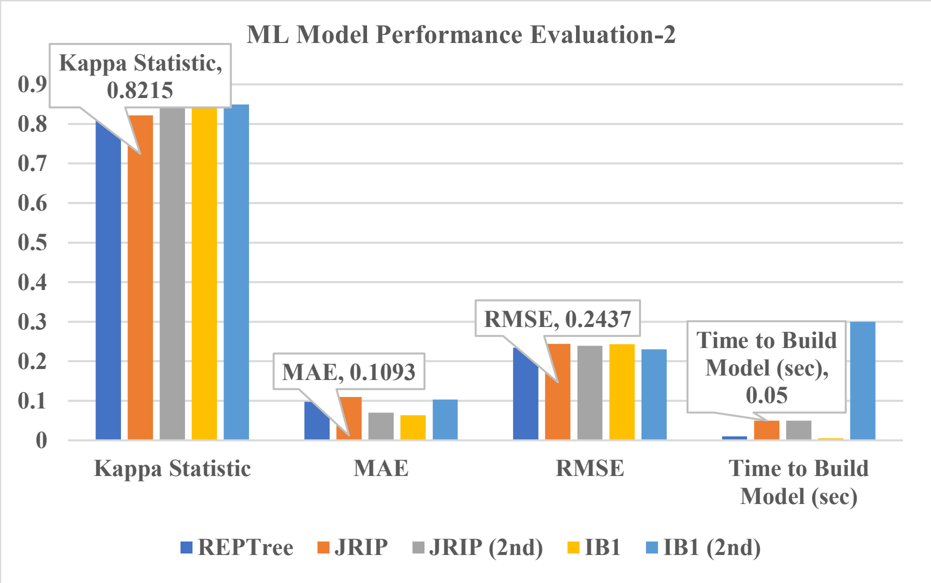

Fig. 6. Model attributes Vs output 2

JRIP (2nd) exhibits a visible enhancement over its novel form, especially in Kappa and MAE, Fig. 6 illustrates how the IBk1 group exhibit error metrics and Kappa, making it highly perfect model for applications requiring high precision and low errors. With its modest Kappa and notable computing time of 0.01 seconds, REPTree offers stable performance, making it a great option for time-sensitive capabilities. IBk1 variants exhibit a strong trade-off between speed and precision, with slower models attaining higher precision while faster models may expense some accuracy.

Table 3 is a Proportional analysis of the confusion matrices from all selected classifiers directing on how well each model classified the 3 sleep disorder categories: None, Sleep Apnea, and Insomnia.

All classifiers performance evaluation, with REPTree and IB1 (2nd) attaining the maximum true positives at 210 and recording only nine misclassifications, while JRIP showed somewhat more confusion, misclassifying six instances as Insomnia. As shown in table 3. JRIP delivered the best performance with 69 correct predictions and only 9 misclassifications, while JRIP (2nd) and IB1 showed slightly more confusion by misclassifying 7 instances as None; meanwhile, REPTree and IB1 (2nd) demonstrated balanced performance with 67 correct predictions.

Table 3. Confusion matrix comparison

Classifier	None (a)	Sleep Apnea (b)	Insomnia (c)
REPTree	210 / 5 / 4	5 / 67 / 6	8 / 6 / 63
JRIP	208 / 5 / 6	5 / 69 / 4	8 / 10 / 59
JRIP (2nd)	209 / 7 / 3	7 / 67 / 4	8 / 3 / 66
IB1	208 / 7 / 4	7 / 67 / 4	7 / 5 / 65
IB1 (2nd)	210 / 5 / 4	5 / 67 / 6	7 / 5 / 65

For the Insomnia class, JRIP (2nd) achieved the highest true positives with 66 and recorded the lowest misclassifications, while JRIP showed slightly more difficulty by misclassifying 10 instances as Sleep Apnea; meanwhile, IB1 (2nd) and REPTree delivered comparable results with 65 and 63 correct predictions respectively.

JRIP (2nd run) and IB1 (2nd run) delivered the most balanced and precise classification across all three classes, while REPTree stood out for its speed and strong performance in the "None" class, though it exhibited slightly higher confusion when differentiating Sleep Apnea and Insomnia. JRIP, as a rule-based and interpretable model, provided transparency but showed greater misclassification in Insomnia during its first run, whereas IB1, being instance-based, maintained consistency and efficiency with minimal training requirements, making it particularly well-suited for real-time or resource-constrained applications.

6 Conclusion and Future scope

Although it takes longer to execute, the IBk1 (2nd) model is the most dependable choice due to its strong RMSE, low error rates, and superior Kappa; JRIP (2nd) is a competitive alternative with balanced performance; and REPTree provides a workable solution for quick deployment with respectable accuracy. By incorporating ensemble approaches, refining hyperparameters for quicker execution, and customizing them for domain-specific datasets like IoT-based monitoring or healthcare diagnostics, these models can be further improved. Furthermore, investigating hybrid approaches that combine the effectiveness of instance-based techniques like IBk1 with the interpretability of rule-based models like JRIP may result in even more reliable and scalable solutions for real-time and high-accuracy applications.

A revolutionary step toward precision healthcare is the incorporation of IoT technologies into sleep analytics, which allows for ongoing, non-invasive, and customized sleep pattern monitoring. IoT systems produce rich streams of physiological and lifestyle data by utilizing wearable technology and ambient sensors. These streams are then processed by machine learning models to accurately classify sleep problems including insomnia and sleep apnea. This synergy facilitates scalable, real-time therapies that are customized to each patient's needs in addition to improving early detection and diagnostic reliability. IoT-driven sleep analytics have enormous potential to connect clinical practice with intelligent healthcare settings as they develop, opening the door to proactive sleep health management and enhanced general wellbeing.

References

1. Karna, B., Sankari, A., Tatikonda, G.: Sleep disorder. In: StatPearls [Internet], StatPearls Publishing, Treasure Island (2025)
2. Gauld, C., Lopez, R., Geoffroy, P.A., Morin, C.M., Guichard, K., Giroux, É., Dauvilliers, Y., Dumas, G., Philip, P., Micoulaud-Franchi, J.A.: A systematic analysis of ICSD-3 diagnostic criteria and proposal for further structured iteration. In: Sleep Medicine Reviews, vol. 58, p. 101439 (2021)
3. Huyett, P., Siegel, N., Bhattacharyya, N.: Prevalence of sleep disorders and association with mortality: Results from the NHANES 2009–2010. In: Laryngoscope, vol. 131, no. 3, pp. 686–689 (2021)
4. Zonayed, M., Tasnim, R., Jhara, S.S., Mimona, M.A., Hussein, M.R., Mobarak, M.H., Salma, U.: Machine learning and IoT in healthcare: Recent advancements, challenges and future direction. In: Advances in Biomarker Sciences and Technology, vol. 7, pp. 335–364 (2025)

5. Dhruba, A.R., Alam, K.N., Khan, M.S., Bourouis, S., Khan, M.M.: Development of an IoT-based sleep apnea monitoring system for healthcare applications. In: Computational and Mathematical Methods in Medicine, article 7152576 (2021)
6. Balaji, A., Sathyasri, B., S, V.V.R., Indumathy, D., Krishnan, R., Vanaja, S.: Intruder Alert System in Smart Home based on IoT Technique. (2022). https://doi.org/10.1109/icpects56089.2022.10047243.
7. Zan, X., Wang, D., Song, C., Liu, F., Xian, X., Berry, R.: Weakly supervised deep learning for monitoring sleep apnea severity using coarse-grained labels. In: IEEE Transactions on Automation Science and Engineering, vol. 22, pp. 15227–15240 (2025)
8. Sinthia, P., M, Malathi., T, Sripriya., Krishnan, R., G, Gurumoorthy., Jalaldeen, K.: Monitoring vital parameters of comatose patients using smart sensors integrated with cloud storage. (2024). https://doi.org/10.1109/i-smac61858.2024.10714845.
9. Ye, G. et al.: Learning unified model for sleep health monitoring. In: Proceedings of 8th International Conference on Communication and Information Systems (ICCIS), pp. 27–32 (2024)
10. Oh, S., Kweon, Y.S., Shin, G.H., Lee, S.W.: MEDi-SOL: Multi ensemble distribution model for estimating sleep onset latency. In: IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 7, pp. 4249–4259 (2024)
11. Vanitha, V., Joe, S.B., Krishnan, R., Fletcher, A.S.A., Anju, M., Akila, V.: Cognitive Threats Detection Model using Nature Inspired Chimpanzee Optimization for IoT Networks (CCM-COM). In: Atlantis highlights in engineering/Atlantis Highlights in Engineering. pp. 629–637 (2025). https://doi.org/10.2991/978-94-6463-754-0_55.
12. Lan, Z., Minhui, D., Yi, Z., Lingyu, S.: Intelligent automatic sleep staging model based on CNN and LSTM. In: Frontiers in Public Health, vol. 10 (2022)
13. Abolfathi, Y., Mohebbi, M.: Heightened heartbeat evoked potential in obstructive sleep apnea disorder during sleep. In: IEEE Access, vol. 12, pp. 189153–189162 (2024)
14. Kalaiselvi, B., Karthik, B., Kumaravel, A.: Predicting level measurements by supervised learning based on GABOR and SMOTE filters: An industrial non-interacting tanks scenario. In: International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 31, pp. 165–179 (2023)
15. Kalaiselvi, B., Karthik, B., Vijayan, T.: IoT framework for measurement and precision agriculture: Predicting the crop using machine learning algorithms. In: Technologies (2022)