Addressing Behavioural Patterns of Late-Night Sleepers Using a Supervised Learning Approach

By
Neloy Pramanik Supto
ID: 213-15-4349

Rajat Chowdhury
ID: 213-15-4271

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised by

Mushfiqur Rahman
Assistant Professor
Department of Computer Science and Engineering Daffodil International University

Co-Supervised by

Dr. Naznin Sultana
Associate Professor
Department of Computer Science and Engineering Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

September 16, 2025

APPROVAL

This Project titled "Addressing Behavioral Patterns of Late-Night Sleepers Using Supervised Learning Approach," submitted by Neloy Pramanik Supto and Rajat Chowdhury, ID NO: - 213-15-4349 & 213-15-4271, to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 16-09-2025.

BOARD OF EXAMINERS

Dr. Arif Mahmud (AM)
Board Chairman
Associate Professor & Associate Head,
Department of CSE, FSIT
Daffodil International University

Dr. Md. Ali Hossain (MAH)
Internal Examiner 1
Associate Professor, Internal Member,
Department of CSE, FSIT
Daffodil International University

Mr. Amir Sohel (ARS)
Internal Examiner 2
Assistant Professor, Internal Member,
Department of CSE, FSIT
Daffodil International University

Dr. Md. Manowarul Islam (DMMI)
External Examiner
Professor, External Member,
Department of CSE, FSIT
Jagannath University

DECLARATION

We hereby declare that this project has been done by us under the supervision of Mushfiqur Rahman, Assistant Professor, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Mushfiqur Rahman (MUR)
Assistant Professor
Department of Computer Science and
Engineering Daffodil International University

Co-Supervised by:

Sultana

Dr. Naznin Sultana (NS)
Associate Professor
Department of Computer Science and
Engineering Daffodil International University

Submitted by:

Neloy 15.09.2025

Neloy Pramanik Supto
Student ID: 213-15-4349
Department of Computer Science and
Engineering
Daffodil International University

RAJAT 15.09.2025

Rajat Chowdhury
Student ID: 213-15-4271
Department of Computer Science and
Engineering
Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Mushfiqur Rahman, Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Machine Learning (ML)** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.
We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Issues related to Sleep have become a big problem among all age Groups, As It leads to long-term physical and mental health problems. This research aims to classify medical conditions according to sleep patterns and lifestyle behaviors associated with them using machine learning algorithms. A dataset was collected that included sleep habits, psychological symptoms, side effects and techniques for managing them. Then, the data was further enriched via SMOTE in order to solve the class imbalance problem. The feature selection algorithm that was used was called Mutual Information. Using Random Forest as a classifier model achieved the highest accuracy of 97.7%, and next is XGBoost at 96.44%, followed closely by Decision Tree with an accuracy of 96.03% and K-Nearest Neighbors with an accuracy of only 83%. Methods that were used also investigated pathways to late sleep and coping mechanisms. It is believed that this research provides an example of the promising scope of supervised machine learning models in health prediction respect of sleep factors and will be useful for people and patients to detect long-term sleep related disorders using data-driven messages.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

This chapter gives you a short and sweet demonstration of the report; from its motivation and the reason for its performance. It readies the reader for the rest of the report. The purpose of the This section presents the research questions, predictions and content of the report to help structure the reader.

1.1 Introduction

In the past few decades, sleep has been recognized as a major area for research due to its impact on overall physical health, cognitive function and educational outcomes. In fact, college students are especially vulnerable to sleep issues because they have irregular daily routines and are pulled in many directions with academic demands as well as overuse of digital devices [13]. Although the older diagnostic methods such as Polysomnography (PSG) remain to be very accurate, they have limitations of being resource consuming in terms of cost and time and may not be scalable for mass screening [24]. ML methods are being employed more and more for evaluating sleep metrics based upon behavioral and physiological measures to cope up with these limitations. Various supervised learning algorithms were employed in the previous works, such as Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest and Extreme Gradient Boosting (XGBoost) which some of them also implemented in this study [27] [25]. The utility of these algorithms has been demonstrated on questionnaire data and physiological signals such as heart rate variability along with EEG features [8]. In the current study, The aim was to propose a novel classification framework for grouping students in various medical conditions according to their delayed sleep practices and correlated factors. The target variable being derived is Medical Condition Category, which contains: Mental Health Issues, Health Issues, Sleep/Respiratory Disorders and No Condition as well as Others. Previous works have also suggested similar multi-class classification frameworks for discernment among insomnia, sleep apnea and healthy subjects [6]. The best performance was obtained in a Random Forest classifier (97.4% of accuracy). This is in line with comparable performance seen in earlier works involving tree-based ensemble methods on lifestyle and clinical datasets for prediction of sleep disorders [3]. Moreover, both support vector machine (SVM) and logistic regression models were used to generate baselines, confirming results from other non-sensor-based detection in the literature. This is in contrast to many other sensor-based models using EEG or actigraphy but rather, questionnaire-based data (including late sleep reasons, physical complaints and ways of dealing with them). This method will allow the installation to be simplified in environments with limited resources like universities dormitories [12].

These publicly available insights from this model (eg, what type of medical group trends most highly for reasons for not sleeping or side effects) can be selectively applied to build personalized behavioral health interventions. Earlier studies have shown that irregular sleep timing, even more so than sleep duration, are associated with worse cognitive performance and circadian health [14]. Hence, this research provides a potentially affordable and scalable alternative for sleep health surveillance in university populations by offering a classical machine learning-behavioral analysis composite. This direction is in line with the changes in scope addressed on digital sleep health and behavioral data analytics [21] .

1.2 Motivation

Proper and enough sleep is a necessary part of life, it affects physical health, mental strength & daily activities. Sleep disorders in terms of reduced quality or irregular, inadequate duration are strongly associated with metabolic disorders, cardiovascular complications, cognitive impairment, and emotional disturbance. However, due to the high cost, complexity and inaccessibility of traditional clinical procedures, real-time large-scale sleep assessment is rare. To fill this gap, in this study we have introduced the development of an individual level machine learning based system that can predict sleep related health status only using behavior and lifestyle features which are more easy to obtain. The outcome variable, Medical Condition Category, dichotomizes individuals into no conditions, comorbidities (yes/no), stigma (yes/no), and sleep or respiratory (yes/no), For precise characterization without the use of invasive diagnostic testing. Based on this predictive capability, the system adopts a machine learning technique in order to provide tailored suggestions in the form of practical and situational tips (possible causes for late SO/SF, expected results, adaptive coping suggestions) and the cluster specific typical sleep to bedtime and wake-up schedules. The advice is then smoothened by the advice logic based on the particular user BMI, thereby making the advice more pertinent. The ability to pair this predictive component with actionable guidance is a real advantage — it enables the model to not only make high-level predictions that are very abstract in terms of classification, but concrete and substantive in terms of language about which an individual can prompt their efforts to insure sleep health. It could potentially offer a scalable, non-stigmatizing and cost-efficient way to raise awareness about the consequences of sleeplessness for health, offer cues for early detection, and fuel healthy sleep behaviors, especially among high risk populations such as university students.

1.3 Rational of the Study

The reason of researches to predict and recommend personalized advice of sleep health are required because sleep problems are increasing compared to various human well-beings. Although the consequences of developing sleep problems such as difficulty sleeping, inconsistent sleep timing, and poor quality rest have been linked to harmful physical, mental, and cognitive effects with consequences that are preventable when detected early, since they have not shown any symptoms of vulnerability, many people do not appear to know their own vulnerability before symptoms suddenly become severe. Conventional diagnostic methods such as polysomnography, while accurate, are expensive and time-consuming, and do not lend themselves to large-scale or continuous monitoring. The net result is a critical demand for scalable, non-invasive, and cost effective early detection methods that provide actionable guidance. For this purpose, to go beyond a dichotomous recommendation, the current work involves a tight coupling between the P4-medical approach, and a machine learning-driven classification. But in to predicting target variable Medical Condition Category (no condition, health problems, mental health problems, sleep/respiratory problems and other) system provide correct risk stratification for user's data present in system. This predictive function is augmented by a recommendation engine that sits on top of this it produces smart recommendations such as why sleep might be disturbed, predicted health impact, recommended management and what you should normally be sleeping at and waking up but improved by adjusting for the user's BMI to make recommendations for you. And not merely building a predictive model, but building the infrastructure all the way up to detection and down to individualized intervention. The type of integration being carried out is such that outcomes are turned from high-level categories into actual user advice to encourage change, while still remaining in a form that's relatively simple to infer with "almost no knowledge." This research will try to bridge the gap between the proved potential of technology and the problem to be solved with classification accuracy and user tailored context aware recommendation. Ultimately, we hope this research will make a contribution to the scientific literature on data- driven sleep health modeling & practice, as well as providing a generally, novel direction toward improving behavior and reducing long term health consequences using the platform that serves daily ads that reach approx 100 million people per week.

1.4 Research Questions

The research work has following main objectives to address in the field of prediction of sleep health using machine learning and personalized recomendation. These are questions that serve as guideposts for a comprehensive investigation of the ways in which the inclusion of computational methods could improve the early detection of sleep disorders and personalized recommendations for optimal general health status. The overarching research questions include:

Q1: From available sleep and lifestyle data, how machine learning models can be used to correctly categorize an individual in one of various medical condition categories?

Ethically motivated and more exciting, as it improves health science, this question attends the question of categorizing individuals in no condition / general health / mental health / sleep respiratory or other. The aim is to investigate to what extent these models can be further pushed and how much (and what kind of) practical large scale tasks they can (efficiently) outperform "standard" approaches on.

Q2: What is the comparative performance of the different machine learning models when predicting the medical condition categories and how much does this influence the overall system performance?

Thus, the issues in this question are about comparison between various algorithms and choosing some models that our system can predict sleep health well with a high accuracy, stability, and robustness. These results may guide decision-makers on which models to implement in eventual use.

Q3: How can such risk scores be linked to personalized sleep recommendations, and how would they differ according to personal characteristics such as the body mass index?

It assesses predictive result (relational integrated outputs to a smart recommender engine) Export (diagram), c Monitoring the factors of why the sleep is really disturbed – and what possible health impact has – an individual coping strategies with individual times of every sleep/wake coordination determines.

1.5 Expected Output

Considered practical implications of this work include evidence for both theory/game refinement and implementation to facilitate scientific progress and new technological possibilities in the space of assessing and personalizing sleep health. The official expectations include:

Towards a Smart System for Predicting and Suggesting Sleep Health for the Development

The end goal of this study, will most certainly be predictive algorithms capable of categorizing people into an existing condition or other (however many choices there are) and doing so with the help of sleep/life style data and some advanced machine learning. It will have predictive modeling and where recommendation module would tell a person happening what his/her sleep should look like, what are reasons for disturbance in the sleep sleep, its influence on health and the perhaps solution the lot depending on tremendous lot of solutions one has considering his/her BMI.

Model Evaluation and Comparison in Machine Learning

Comprehensive analysis and comparison on selected machine learning methods. This should make this info be reported for the other performance metrics (accuracy, precision, recall, f1 score and whatever more robustness stuff is out there). The results of these studies will provide evidence to determine the optimal approach from an algorithm perspective (including scalability and reliability) to estimate sleep health.

Personalised Recommendations Techniques Integration

We expect the presented approach that combines predicted outputs with a knowledge-structured recommendation engine to offer meaningful and personalised sleep recommendations. This in turn would have the potential to promote high levels of user engagement, behaviour change and preventative health through data driven information.

Methodology Data processing, feature engineering techniques applied

It will also make use of more advanced data preprocessing, feature selection and transformation processes to improve the model. These latter methods are anticipated to improve the modeling prediction ability to handle mixed-data from several populations or wearables.

Test rigorously: validate results

Furthermore, such a study will use advanced data pre-processing, feature (gene) selection and transformation methods to improve the model. Such approaches are anticipated to improve the prediction accuracy of the model while being able to perform on mixed-data that originates from different populations or wearables.

Contribution to Academic and Practitioners' Literatures

The research is intended to be innovative; machine learning, digital health and sleep research and its legacy will last. Ultimately, we anticipate that our findings will also inform the development of technology-based sleep health interventions with health outcomes as the end target (i.e., to enhance personal wellness and impact population health).

1.6 Organization of the Report

This draft report is arranged in a series of chapters that are meant to guide the reader through the research project, its results and their possible implications. To ensure coherence, clarity, and depth to the instructional design each chapter has been moulded around specific aspects of this study. Now, the chapter wise intro discussion is shown here:

Chapter 1: Introduction

In this chapter, the research is presented, and its importance in the prediction of sleep health and provision of personalized suggestions with machine learning algorithms is argued for. The emphasis on how sleep affects general health and academic or professional life makes clear the necessity for an automated, data-driven solution. This set the basis for presenting the research objectives, questions, and herein discussed conceptual framework.

Chapter 2: Background

In this chapter, the research is presented, and its importance in the prediction of sleep health and provision of personalized suggestions with machine learning algorithms is argued for. The emphasis on how sleep affects general health and academic or professional life makes clear the necessity for an automated, data-driven solution. This set the basis for presenting the research objectives, questions, and herein discussed conceptual framework.

Chapter 3: Research Methodology

This chapter explains the methodology followed in the study. It lists the dataset selection, preprocessing steps, definition of target variable and then feature extraction methods. A detailed description of how the machine learning models have been selected are Additionally, the description of the selection of machine learning models are designed and configured, and the evaluation metrics utilized for measuring performance are explained. It also raises issues of ethical data privacy and informed consent.

Chapter 4: Experimental Result

Predictions of the model, a learning rate, if prominent, are presented in the experimental results of the next chapter, with sleep health, and personalized recommendations in sleep – all as new ground of the predictors. Supported The detailed comparative results of various algorithms are included: accuracy scores, precision-recall metrics, and visual plots. A systematic effort is made to study how the accuracy of the models varies with different feature sets or configurations.

Chapter 5: Impact on Society

This chapter is the first step of inspection to test whether or not it is socially acceptable to fix the sleep health prediction system by automated systems and monitors the society. Applications to healthcare/workplace wellness/academic performance improvement/control application for personal health management and commercial lease tenant power consumption applications are discussed. The objective of the study is to assess the consequences of providing individualized, tailored sleep recommendations to people; this could potentially lead to ameliorating the social and personal health implications

This chapter is the first step of inspection to test whether or not it is socially acceptable to fix the sleep health prediction system by automated systems and monitors the society. Applications to healthcare/workplace wellness/academic performance improvement/control application for personal health management and commercial lease tenant power consumption applications are discussed. The objective of the study is to assess the consequences of providing individualized, tailored sleep recommendations to people; this could potentially lead to ameliorating the social and personal health implications

Chapter 6: Summary, Conclusion, Recommendation, and Implications for Future Research

The final chapter of this thesis will summarise the key findings and describe the impact on the prediction of sleep health. Luckily deployment recommendations can be had and some considerations for consumer facing use are mentioned. The limitations of this study have been recognized and some possible directions for future research that could ameliorate prediction performance, model generalizability, and the possibilities for model personalization are proposed. This hierarchical structure facilitates a rational progression of the initial conceptualisation of the research problem, to the findings where results are discussed with consideration of public good and limitation/next steps. The system has been designed so the user clearly can read both the scientific influence and the practical implications as they apply to the study of sleep health, to express the academically rich view.

Chapter 2

Background

The background of the study, starting with important preliminaries and terminologies that are used in the present research, is provided in this chapter. The literature review is significantly contributing to the development of research questions or hypotheses and is part of the understanding and synthesis of literature about a study topic. The chapter concludes with a summary of the findings and discussions.

2.1 Introduction

Sleep is a key element essential to physical, mental, and social health. Poor sleep or sleeping habits can result in severe health issues and is also a cause of heart diseases, diabetes, obesity and depression. Sleep health has been of increasing interest to researchers in recent years due to its long-term association with mental and physical conditions. Nevertheless, the conventional methods for the diagnosis of sleep disorders such as PSG are expensive and time consuming and are not feasible for large or continuous monitoring. THIS is why machine learning methods have been increasingly used because they can analyze lifestyle and behavioral data to categorize sleep related health issues more accurately. Chapter 2 provides the background of the study by reviewing related literatures, similar applications, research gaps and the summary of the research problem.

2.2 Preliminaries/Terminologies

To have a smooth and consistent nomenclature throughout the present study, certain essential concepts and definitions that are necessary to understand the present research are described. This is the subset extracted from the well-organized CSV files that I have which included demographic information, lifestyle, and sleep habits. Target variable to be predicted for the individual would be (from Health Issues, Mental Health Issues, No Condition, Others, Sleep/Respiratory Disorders) as Medical condition category (5 categories). For predictive modeling, you would like to be aware of that. From a data perspective, dat a preprocessing refers to a wide variety of data preparation tasks for raw data to be more readable. In this study, there are the missing value processing of each descriptor, the conversion of their categorical labels into numerical representations, one-hot encoding of the multiselect values, and a statistical weak feature selection of a few critical descriptors (e.g., mutual information). Class Imbalance: We use a variety of over-sampling techniques such as SMOTE to ensure an equal representation for all health classes when training the model. Here, the machine learning (ML) relies on algorithms that are capable of capturing the complex patterns and relationships among the inputs and the target at a time. The submission is based on predictive analysis using models (Random Forest, Gradient Boosting, Logistic Regression and Support Vector Machines). We chose these algorithms as they have demonstrated good performance on heterogenous tabular input data and are highly interpretable with decent accuracy. The evaluation metrics in this work are accuracy, precision, recall and F1-score with which the performance of each classifier is estimated. The competency of the results is guaranteed by use of cross-validation methods. Recommendations are made by recommending feedback received after classification. The system then retrieves prerecorded advice from the classified medical condition (e.g., work out what is a normal bedtime, diurnal sleep/wake, sleep/wake outside normal window). By mapping the process the above takes domain knowledge and statistical patterns in the data and provides you with actionable sleep habit recommendations unique to you. For the purpose of the analysis, given that these terms are used with their traditional definition in machine learning and health data analysis, we here use them in order to preserve the correctness of our description of the methods and of the results of our work.

2.3 Literature Reviews

Literature on the topic indicates an explosion of machine learning (ML) application within the sleep research spectrum — from sleep staging and disorder detection, to behavioural health modelling and genetics analysis. These investigations employ a broad range of data, including EEG, heart rate variability (HRV) and wrist accelerometry and even genomic data. This section provides a summary of major contributions along keylines, it highlights performance oriented approaches and emerging largescale deployments. A Machine learning approach to optimize the sleep schedules of college students has been presented by [1]. We obtained 300+ hours of sleep data for one subject, which was used as labeled input to predict bed and wake times and total sleep time with SN algorithms -SVM, KNN, LDA, DTC. Of which, SVM performed out to be the best-model reaching an accuracy of more than 65%. This model was then used to compute the optimal schedule, which applied a bedtime at 01:00 AM and wake time at 10:00 AM (with a total sleep duration of 9.5 hours). However, the study is only based on data from a single participant and is limited by some practical constraints of using Muse headset and thus future work should extend these findings. The authors also highlighted the necessity in future works to employ larger and more diverse datasets, recruit more participants, and use the raw EEG data for a better performance of models as well as generalization.

In another research, an advanced neural network (NN)-based framework for sleep stage analysis and narcolepsy diagnosis was introduced by [2]. They pooled data from 10 cohorts performed at 12 major sleep centers on different continents to achieve a large base of patients. There are two modules in the study that study includes one is sleep stage classification the second is biomarker development for T1N. For the classification of sleep stages the accuracy of Convolutional and Recurrent neural networks (CNN and RNN) was evaluated in combination with various features of REM-latency, SOREMPs, and NREM-fragmentation. They used the GP models for better diagnostic decisions that often involve uncertainty that is critical for deciding on treatment for a patient. Besides, HLA-DQB1*06:02 gene test in combination improved the specificity. Various ensemble modeling methods cooperated from distinct perspectives on classification or regression tasks and dramatically strengthened model robustness and accuracy. These findings highlight that consideration should be given to models such as these for large-scale use not only in faster and more accurate identification of narcolepsy, but also in extensions of home sleep disorder diagnostics of narcolepsy.

A hybrid of feature selection and classification model was presented by [3], where the sleep disorders can be predicted by agriculturally foraged food parts. This was achieved using the Binary Dipper Throated Optimization (bDTO) method to select the most influence features from the 400 health and lifestyle data case study. They later applied Logistic Regression on the data and achieved 95%. bDTO resulted in a reduced level of error rates than in other methods and Logistic Regression classifier showed the best performance for the progression of early disease diagnostic tool in to the personalized sleep disorder screening. The power and advantage of linking bio-inspired optimization algorithms with traditional statistical models in the context of sleep health research, as shown in our study.

An alternative to EEG-based polysomnography for sleep stage classification was introduced by [2], where heart-rate variability (HRV) signals were utilized in combination with Long Short-Term Memory (LSTM) neural networks. In this article 292 subjects were used from 584 nights to train and test. When tested the proposed system produced an accuracy of 77%, with a Cohen's κ of 0.61 and outperformed existing HRV based classifications. Moreover, the work presented in this paper demonstrated that LSTM networks could be used for the accurate long-term modelling of temporal associations inherent to the physiological signals. But slowed performance was observed in normal elderly and individuals who had Parkinson's disease, suggesting generalisability issues in a heterogeneous population.

In the present work, [5] developed sleep stage classification technique using ELM with PSO. It used the input of the ECG derived HRV signals and it applied the PSO not only for feature selection, but also for the optimization of the hidden nodes in the ELM. Tested on MIT-BIH dataset, it reached 82.1% accuracy in binary classification and outperformed ELM and SVM in generalization power and computational requirement. The research really showcases how powerful hybrid machine learning models can be for sleep research and big data healthcare. The proposed work In [6], a machine learning model was implemented to perform the automatic classification among the three types of sleep disorders, separating the normal cases from insomnia and sleep apnea. Three classification approaches were investigated: OVA: One vs All and OVO:

One vs One both with Logistic Regression (LR) and Support Vector Machine (SVM) classification methods over a dataset that included 374 instances with health, lifestyle, and sleep features. The criterion for splitting was Gini impurity and information gain. Best performing SVM model was polynomial SVM degree two with 91.44% Accuracy and false positive rate 0.24 per encounter, Cohen's κ = 84.97%. These are the obvious forerunners in question-based, non-invasive early detection techniques for sleep disorders.

Focus specifically on a more comprehensive review of applications in machine learning for behavioral health, and have instead reviewed the article by [7] on PA, sedentary behaviour and sleep behaviors through wearable data sources. A survey was conducted to review supervised and unsupervised learning type algorithms used in behavior classification, profiling and para-dictive modeling. The review also identified obstacles to model validation, interpretability, and feasibility of application, yet it is noted that machine learning afforded a unique opportunity to improve behavioral health research. While promising, a study has said the research is limited in its impact by both methodological challenges and concerns around transparency. A paper introduced a supervised machine learning model to detect sleep bruxism from EEG signals based on non-linear dynamic features (NLDF). PhysioNet CAP sleep data were preprocessed by segmenting and filtering recordings in different frequency bands and sleep stages. Beta and gamma activities were featured extracted using Katz fractal dimension and Lyapunov exponent. Of all the classifiers used, SVM with RBF Kernel accomplished the best accuracy = 93.5% The results indicated the diagnostic value of non-linear EEG properties in automatic sleep disorder detection [8].

Another Study, conducted a systematic review of the literature to evaluate the link between sleep timing, sleep stability and health outcomes in adults. Based on evidence from 41 studies including a total of over 92,000 participants, the review showed that people with later and irregular sleep were more likely to experience adverse outcomes, including poor mental health, an increased risk of type 2 diabetes and lower cognitive function. Weekend catch-up sleep effects were also reported. The overall evidence quality was from very low to moderate, supporting public health messages for earlier and more stable sleep timing on subjective well-being [9]. In recent research, presented a machine learning approach to predict sleep quality using HRV and facial skin temperature recorded during wakefulness. Using a dataset of 28 adults, the best performance was achieved with Support Vector Machines (SVMs) which reached an accuracy of 83.4% when combining HRV and IRT features. It demonstrated a non-invasive, inexpensive but powerful way with real-world application for workplace monitoring and health assessment using a wearable contact-less sensing system [10]. A grid of force sensing resistors (FSRs) was placed under the mattress for sleep behavior analysis in real-life dataset. In this research, the classification of sleep positions (e.g., supine, prone, lateral) is compared using various machine learning algorithms, such as Random Forest, k-Nearest Neighbors (IbK), and Logistic Regression. The Random Forest is the winner with 95.7%. The system has potential value for long-term sleep monitoring that could be useful in pressure sore prevention and home care [11]. A paper presented a plethora of computational methods, as they reviewed the sleep behavior analysis. The methods employed in the review look diverse across polysomnography, actigraphy, wearable/nonwearable sensor systems, computational tools by machine learning (ML), deep learning for sleep stage classification, posture recognition, disorder detection. The technology was placed within its current dimensions of life in the clinic and home, by discussing pluses and minuses of different methods [12]. Okano et al. explored the relationship between sleep and academic performance among 88 college students at MIT, utilizing Fitbit wearables to acquire objective measurements of sleep over the course of a semester. As a result, longer, better quality, and more consistent sleep were all significantly predictive of better academic performance, accounting for a total of 24% of the variance in grades. Interestingly, sleep in night before examinations did not predict outcomes; instead, cumulative sleep patterns over preceding weeks or months were more influential. Gender based differences were also observed, as female students demonstrated higher sleep quality and more consistent schedules, which partly accounted for better academic performance [13]. In a study, sleep regularity was also found to influence academic performance and circadian timing in 61 undergraduates who were continuously monitored for 30 days. Introduction: A novel Sleep Regularity Index (SRI) captures the stability of sleep–wake cycles. Irregular sleepers also had a delay in circadian alignment increased daytime sleep and lower academic GPA. Most of these differences were due to abnormalities in patterns of exposure to light rather than inherent changes in circadian behavior. The authors of the study concluded that increased sleep regularity might improve not only academic performance, but also circadian stability [14].

In another work, which investigated machine learning approaches for sleep disorder prediction that compared about 23 regression and 29 classification models on mammal-based datasets. In regression, MLP and Ibk produced the best results overall; in classification, SMOreg and Random Forest. The Function learning strategy showed the highest average performance of all the strategies, comprising three super-strategies. Ensemble methods taking advantage of these models may increase performance and the generalizability for real-world use was suggested using light-weight, sensor-based approaches [15].

In a review of pathways through which sleep disorders and inadequate sleep contribute to activity in the sympathetic nervous system, one factor is known to be an important promoter of cardiovascular disease [16]. In particular, the majority of patients who are sleep deprived, and those with insomnia, narcolepsy or obstructive sleep apnea often have increased sympathetic responses accurately quantified by microneurography and plasma norepinephrine measurements. Results highlighted sex differences, indicating that women may be particularly vulnerable to dysregulation and related cardiovascular risks. The review highlighted the need for future balanced sex representation and direct non-hop efficacy testing.

A paper conducted a genome-wide association study (GWAS) utilizing accelerometer data from 85,670 participants in the UK Biobank to examine the genetic determinants of sleep characteristics. Examinations of eight metrics concerning sleep duration, quality, and timing revealed 47 significant genetic associations, comprising 26 novel loci. Significant findings comprised a PDE11A missense variant associated with sleep duration and efficiency, alongside the enrichment of serotonin-related genes affecting sleep quality. Genetic correlations between accelerometer-derived and self-reported measures were modest; however, the study offered novel insights into the biological mechanisms influencing sleep behavior [17]. Another paper performed a systematic review and dose–response meta-analysis of 67 prospective cohort studies, encompassing over 3.5 million participants, to investigate the correlation between sleep duration and health outcomes. A U-shaped correlation was observed, indicating that both insufficient sleep (<7 hours) and excessive sleep (>7 hours) are linked to heightened risks of all-cause mortality and cardiovascular incidents. The risk was lowest at about seven hours of sleep each night. Every hour less than seven raised the risk by about 6%, and every hour more than seven raised the risk by up to 18%. These findings underscored the essential significance of sustaining optimal sleep duration for enduring cardiovascular health [18].

An exciting paper investigated the use of Random Forest models for sleep classification and nonwear detection from accelerometer data on the wrist. A cohort of the 158 individuals, including normal and sleep disordered subjects, was used for the training and testing. Its F1-score is 73.93%, and it outperforms all heuristic-based methods. The nonwear detection achieved an F1-score of more than 93%. It was still hard to tell which sleep stage someone was in (REM, N1, N2, N3) because of signal problems, but nap detection was somewhat related to self-reports (r = 0.60). The study offered open-source tools and validated the benefits of Random Forests for extensive accelerometer-based sleep research [19]. The paper developed SleepNet: Self-contemplated deep learning framework for sleep stage detection using a robust accelerometer," hosted on arXiv preprint. This was validated in 1113 polysomnography nights and demonstrated fair agreement in distinguishing sleep–wake states as well as REM/NREM stages. The framework was then applied to accelerometer data in over 66,000 UK Biobank participants, which concluded that habitual short sleep (<6 hours), regardless of efficiency, increased all-cause mortality. This study illustrated that self-supervised methods are capable of achieving scalability in both population-level sleep research and clinical risk assessment by providing an open-source tool [20]. A paper introduced the Digital Sleep Framework through a data-driven approach to change sleep science and medicine. The architectural framework consisted of five modular and inter-related stages: acquiring data (from polysomnography to wearables), storing and curating data (IoT, cloud systems), pre-processing the data (cleaning and transformation), building models (machine learning, deep learning) and applying them in clinical care, consumer health and public health. The paper noted that sensor technology and AI-driven diagnostics present key opportunities and underscored challenges around data standardization, validation, and interpretability. Regarding the development and implementation of digital sleep health tools, the importance of transparency and interdisciplinary collaboration was emphasized by the authors [21].

In a research also use a biLSTM architecture for classifying 5 sleep stages Wake, N1, N2, N3 and REM — based on non-EEG signals. The framework was applied to four physiological signals (heart rate, respiratory rate, cardiorespiratory coupling, and body movement) collected from 123 adults suspected of having sleep disorders. The model showed 71.2% accuracy, $\kappa = 0.43$ and F1-score = 0.65 in average. Performance decreased as the AHI and age increased, and improved with higher sleep efficiency. Correlations with polysomnography findings provided evidence to support the possibility of using non-EEG data for a home-based sleep monitoring device [22].

A model proposed the Sleep Regularity Index (SRI) in a 30-day longitudinal study of 61 Harvard undergraduates to measure daily sleep–wake regularity. Irregular sleepers demonstrated delayed melatonin onset, sleep propensity rhythm, and lower light exposure during the day. Notably, SRI was positively associated with GPA; however, total sleep time was not a predictor. Circadian delays of irregular sleepers were explained primarily by abnormal light exposure, according to modeling analyses. The results highlighted how, regardless of sleep duration, sleep regularity is essential for circadian alignment and school performance [23].

A study examined whether high-resolution physical activity measurements collected during wakefulness via wearable actigraphy were helpful in predicting sleep quality. oth traditional methods (logistic regression) and a variety of deep learning models (CNN, LSTM, and TB-LSTM) were evaluated. TB-LSTM yielded the best performance, including AUC of 0.9714, best recall, and overall accuracy for deep learning approaches that consistently outperformed traditional models. Similarly, CNN is performed extraordinarily well, as its AUC was equal to 0.9456, however logistic regression with 0.6463. Results indicated that deep learning using wearable activity data can produce accurate, non-invasive estimates of sleep quality that can be leveraged for digital health applications and early diagnostics related to sleep quality [24].

An investigated the Personalized machine learning models for estimating sleep parameters from actigraphy data, using polysomnography (PSG) as golden standard. A study made a comparison among generalized models that used population-level data versus personalized models trained with individual-specific data. Compared with the standard approach, a personalized analysis led to substantial improvements in the estimates of important parameters that reflected sleep quality (TST, WASO and SE) and recurrent arousals (NA). Among classifiers, Secure Logistic Regression, Adaptive Boosting and Extreme Gradient Boosting all achieved accuracies of approximately 86%, outperforming Naive Bayes (75%) and about the same as Random Forest (85%). These results indicate that customizing m odling was successful in modeling sleep–wake state prediction [25]. A utilized machine learning classifiers on EEG and EMG recordings in rodents in order to create an automated classification of sleep stages into paradoxical, slow-wave and wake states. The manually annotated recordings and their metadata (complete sets) resulted in 427 hours. The highest accuracy rate was obtained with Random Forest among all of the classification methods (95.78%), which was followed by Artificial Neural Networks with 93.31% accuracy, and the accuracy was lower for Logistic Regression and Naive Bayes. These results suggest that an on-line pre-trained machine learning-based method is feasible and robust for sleep stage determination in animal models and reveals new possibilities for preclinical research [26].

A study in the prediction of obstructive sleep apnea syndrome (OSAS) severity based on demographic, spirometry, gas-exchange and questionnaire data from 313 patients. All machine learning models of our experiment were evaluated for classification and regression. The maximum classification accuracy (approximately 44%) was obtained by Support Vector Machine and Random

Forest models, which was still far from an optimal level for these models to be used as efficient primary diagnostic methods in clinical practice. However, the study shows that machine learning may have a role in initial screening and characterization of features in OSAS-related research [27].

A research performed one of the largest benchmarking studies on sleep–wake classification using the ethnically diverse MESA Sleep dataset that included combined PSG and actigraphy data from 1817 participants. Two tasks were studied: sleep-only classification (from nighttime recordings only) and full 24-hour classification, by comparing outputs from two typical machine learning methods to deep learning methods. These night-models achieved accuracies of 83.1% and 85.4% in the night-task, outperforming all device-based algorithms and rule-based classifiers. Source: Kaggle. At the 24-hour level task, LSTM achieves the highest accuracy of 88.2%, followed by the overall accuracy of CNN models which is 87.7%. These results form a solid baseline for machine learning in large-scale sleep–wake classification [28].

A research team used multi-variable machine learning analysis to explore sleep-wake patterns in DOC patients (12 UWS and 11 MCS patients). An interpretable classifier (Random Forests, RF) and a non-interpretable neural network (NN) were trained on control subjects using a feature robust to noise (permutation entropy) and tested on patients. For 11 patients, the average F1-score score was 0.87 indicating good agreement between predicted and visually checked periods of closing the eye. Taken together, these results suggest increased order in sleep–wake organization across states of consciousness that may reflect MCS vs. UWS at large [29].

2.4 Comparative Analysis and Summary

The comparative analysis of the reviewed literature highlights the diversity of methodologies, datasets, and outcomes in the application of machine learning for sleep research. Traditional approaches such as Support Vector Machines, Logistic Regression, and Random Forest have been widely applied, particularly in studies involving questionnaire data, physiological signals, or wearable-derived activity metrics. While these models often demonstrated moderate to high performance (65–95%), they were generally more interpretable but less adaptable when applied to heterogeneous populations.

In contrast, deep learning frameworks such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), biLSTM, and self-supervised models exhibited superior capacity to capture temporal and nonlinear dependencies in physiological and behavioral signals. These models achieved high accuracies in several studies, with reported performance exceeding 90% in sleep stage classification or disorder detection tasks. Nevertheless, the requirement for large-scale, annotated datasets and the challenge of interpretability were noted as persistent limitations.

The reviewed works also demonstrated variation in data modalities. EEG-based studies provided high diagnostic precision for disorders such as bruxism and narcolepsy but often required invasive or costly measurements. On the other hand, heart-rate variability, accelerometry, and actigraphy emerged as non-invasive, scalable alternatives, though with somewhat reduced accuracy in multi-class classification of sleep stages. Large cohort studies such as the MESA Sleep dataset and the UK Biobank enabled population-scale analyses and supported the development of generalizable models, while smaller participant-focused studies highlighted the benefits of personalized modeling, which consistently outperformed population-based general approaches.

From an application perspective, research extended beyond sleep staging and disorder detection to include associations with academic performance, cardiovascular health, and genetic determinants of sleep behavior. Studies linking irregular sleep patterns and poor academic outcomes, or uncovering genetic loci influencing sleep duration and efficiency, expanded the scope of machine learning in sleep research into behavioral and biomedical domains.

In summary, the comparative analysis reveals that machine learning and deep learning methods have significantly advanced the capacity to classify, predict, and understand sleep-related outcomes. Traditional models remain useful for interpretable screening tasks, whereas deep learning and hybrid optimization-based methods dominate high-performance classification. However, issues of generalizability, dataset imbalance, and real-world clinical integration remain unresolved. These insights underline the need for future studies to bridge methodological rigor with practical applicability, positioning data-driven approaches as essential tools in both clinical and non-clinical sleep research.

Table 2.1: Comparative Analysis Table

Study (Author, Year)	Data/Signal Type	Method/Model(s) Applied	Best Reported Accuracy/Metric	Key Contribution
Azuara-Hernandez & Gillette (2022)	EEG (Muse headset), single subject	SVM, KNN, LDA, DTC	SVM: >65% Accuracy	Optimal schedule estimation for college students
Stephansen et al. (2018)	Multi-cohort PSG, genetics	CNN, RNN, GP, Ensemble	High accuracy; improved specificity with the HLA test	Narcolepsy diagnosis biomarker
El-Kenawy et al. (2022)	Health & lifestyle (400 records)	bDTO + Logistic Regression	95% Accuracy	Feature optimization + classification
Radha et al. (2019)	HRV (292 participants, 584 nights)	LSTM Neural Network	77% Accuracy; κ = 0.61	Low-cost EEG alternative
Surantha et al. (2021)	ECG-HRV (MIT-BIH dataset)	ELM + PSO	82.1% Accuracy	Hybrid ML with improved speed
Dritsas & Trigka (2024)	Health/lifestyle questionnaire	SVM (Polynomial), Logistic Regression	91.44% Accuracy; κ = 0.85	Multi-class disorder screening
Tiwari et al. (2022)	EEG (CAP Sleep Database)	SVM-RBF on NLDF features	93.5% Accuracy	Bruxism disorder detection
Chaput et al. (2020)	41 cohort studies (>92,000 adults)	Systematic Review	Evidence quality: low–moderate	Timing/consistency linked to health
Di Credico et al. (2024)	HRV + skin temperature (28 adults)	SVM	83.4% Accuracy	Wearable/contactless prediction

Study (Author, Year)	Data/Signal Type	Method/Model(s) Applied	Best Reported Accuracy/Metric	Key Contribution
Crivello et al. (2020)	Force Sensing Resistors (FSR)	RF, KNN, Logistic Regression	RF: 95.7% Accuracy	Sleep posture classification
Fallmann & Chen (2019)	Survey across modalities	Review (ML & DL methods)	N/A	Computational sleep behavior survey
Okano et al. (2019)	Fitbit wearables (88 MIT students)	Correlational Analysis	Sleep explained 24% GPA variance	Sleep & academic outcomes
Phillips et al. (2017)	Actigraphy/light exposure (61 students)	Sleep Regularity Index (SRI)	Correlated with GPA	Sleep regularity & circadian alignment
Alazaidah et al. (2024)	Mammal-based datasets	23 regression & 29 classification models	Ensemble; Function strategy	Multi-model comparison for disorder prediction
Greenlund & Carter (2022)	Physiological measures (literature)	Review (microneurography, norepinephrine)	N/A	Sympathetic dysregulation in disorders
Jones et al. (2019)	Accelerometer + GWAS (85,670)	Genetic association	47 associations; 26 novel loci	Genetics of sleep traits
Yin et al. (2017)	67 prospective cohorts (3.5M+)	Meta-analysis	U-shaped risk curve	Sleep duration & mortality
Sundararajan et al. (2021)	Wrist accelerometer (158)	Random Forest	F1-score: 73.93%	Sleep–wake & nonwear detection
Yuan et al. (2024)	Accelerometer (66k Biobank + PSG)	Self-supervised DL (SleepNet)	Fair PSG agreement; mortality link	Scalable population-level sleep staging
Perez-Pozuelo et al. (2020)	Multi-modal data (review)	Digital Sleep Framework	N/A	Data-driven sleep science roadmap
Morokuma et al. (2023)	HR, RR, coupling,	biLSTM	71.2% Accuracy; F1 = 0.65	Non-EEG staging in

Study (Author, Year)	Data/Signal Type	Method/Model(s) Applied	Best Reported Accuracy/Metric	Key Contribution
	movement (123)			suspected disorders
Sathyanarayana et al. (2016)	Wearable actigraphy (adolescents)	CNN, LSTM, TB-LSTM	TB-LSTM: AUC = 0.9714	Predict sleep quality from wake data
Khademi et al. (2019)	Actigraphy + PSG	Personalized ML (RF, RLR, XGB, AB)	~86% Accuracy (personalized)	Personalized vs generalized comparison
Smith et al. (2021)	EEG/EMG (rodents, 427 hrs)	RF, ANN, DT, LR, NB	RF: 95.78% Accuracy	Preclinical sleep classification
Mencar et al. (2020)	Clinical (313 OSAS patients)	SVM, RF, k-NN	~44% Accuracy	OSAS severity prediction (limited)
Palotti et al. (2019)	PSG + actigraphy (MESA, 1817)	CNN, LSTM, Extra Trees	LSTM: 88.2% Accuracy	Benchmark for sleep–wake classification
Wielek et al. (2018)	PSG + video (23 DOC patients)	RF, FFNN, clustering	F1-score: 0.87 (subset)	Sleep organization in DOC patients

2.4.1 Similar Research

Many machine learning applications in sleep research are conceptually similar to the study reported here. A case in point, Random Forest and SVM have been applied in order to classify whether a person has insomnia, sleep apnea or no disorder with accuracies surpassing 90% in some cases. XGBoost has achieved promising results in managing imbalanced sleep data, and it can identify sleep and life style patterns influencing sleep. Other works have used KNN and Logistic Regression for smaller datasets where interpretability was highly valued. Besides disorder detection, machine learning has been applied to predict school performance with sleep data, develop mobile health applications to deliver sleep recommendations and analyze population data sets to investigate risk factors for sleep. “These applications demonstrate, in a tangible way, that if you harness data-driven prediction to provide actionable advice, you can make machine learning really useful in health and everyday life.

2.5 Gap Analysis

Table 2.2: Gap Analysis Table.

Features	Existing Studies	Our Work
Identify lifestyle and medical impact of poor sleep	Yes	Yes
Handle missing and messy data properly	No	Yes
Balance class imbalance in the dataset	No	Yes
Use many models (LR, DT, KNN, RF, XGBoost)	Yes	Yes
High accuracy and generalizable models	No	No (partial improvement)
Explainable results (reason behind prediction)	No	Yes (SHAP, LIME)
Provide user recommendations for sleep improvement	No	Yes
Large and diverse dataset used	No	No (future need)

2.6 Summary

In conclusion, this background chapter has set the scene by presenting the significance of sleep research and demonstrating how machine learning has been used extensively to study sleep health. Prior works show encouraging results in sleep disorder classification, health-risk prediction, and coupling of sleep with academic or cognitive performance. Nevertheless, there are still some challenges in its scalability, dataset imbalancedness, interpretability and personalized recommendations integration. This paper aims to address these deficiencies by customization a su- pervised machine learning pipeline for classifying diseases considering only sleep and lifestyle as inputs as well as providing personalized interventions for improving sleep. This brings high-level computational techniques closer the practical application in health, with the scientific and practical value added to each other.

Chapter 3

Research Methodology

This chapter describes the research methodology from the research object and requirements. It describes the proposed process, functional and non-functional requirements and a data and workflow diagram. The rest of the section provides the details of methodology, the project plan, task distribution and ends with a summary.

3.1 Research Subject and Requirement analysis

The research object of this paper is to solve the scene of machine learning methods in sleep health, specifically to predict diseases related to sleep. Sleep disorders are very common in the general population and the consequences such as cardiovascular morbidity, metabolic disturbances, neurocognitive deficits, and quality of life impairment are well known.

Such conventional techniques, for diagnosing OSA using the gold standard approach (i.e., polysomnography [PSG]), have become increasingly more sophisticated, as well as expensive and invasive, such that they are not feasible for routine use nor durable in large or unceasing monitoring investigations. To enable a more comprehensive inquiry of understanding COPD disease burden, this study intends to carry out machine learning models based on lifestyle and anthropometric indicators and sleep habit data. We have developed a scalable, accessible, and interpretable pipeline for predicting sleep disorders utilizing the capability of these methods. The relative complexity of the study consists in KNN, Logistic Regression, Decision Trees, Random Forest and XGBoost that have been done. After careful preprocessing, feature engineering, and hyperparameter tuning, where Random Forest with best overall performance was emphasised. The research is not limited to just classification, It also added some explainability layers like feature importance and shap/lime analysis to make the model outcomes interpretable.

Secondly, the framework is widened through the introduction of a recommendation module that provides actionable TBL related input. This derived model is then used to identify a list of viable hypotheses supported by experimental evidence and connect them with known causes, side effects profiles and life-style interventions in an interpretable manner for serving as explainable digital sleep health assistant. The research area is at the intersection of machine learning, digital health, and preventive medicine. Both from a technical and a practical point of view there is an emphasis on enhancing predictive performance, making recommendations that users can see. These findings are expected to drive further predictive modeling work in academia and added value propositions for real world healthcare outcomes where early screening allows for personalization of intervention including improving terms of sleep health.

3.1.1 Overview

The data were primarily collected using a structured Google Form. The form link was distributed among diverse participants across various settings, including university campuses, residential halls, urban areas, rural regions, and public streets. In addition to online responses, data were also gathered through in-person interactions and verbal communication with individuals. The information obtained through these conversations was subsequently entered into the Google Form to maintain consistency in data recording. Specific locations in Bangladesh where data collection was conducted include cities such as Rajbari, Kushtia, Sirajgonj, and Dhaka. Participants were also approached in various villages within these cities, as well as at Daffodil International University and its associated residential halls (23.875816, 90.323511). This broad geographic coverage contributed to the richness and variability of the dataset.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2610 entries, 0 to 2609
Data columns (total 21 columns):
 #   Column                                                                  Non-Null Count  Dtype  
---  ------                                                                  --------------  -----  
 0   Your Age                                                                2610 non-null   object 
 1   What is your weight                                                     2610 non-null   float64
 2   Your Height                                                             2610 non-null   float64
 3   What is your gender?                                                    2610 non-null   object 
 4   What is your occupation?                                                2610 non-null   object 
 5   What time do you usually go to bed?                                     2610 non-null   object 
 6   What time do you usually wake up on working days?                       2610 non-null   object 
 7   What time do you usually go to bed on weekends?                         2610 non-null   object 
 8   What time do you usually wake up on weekends?                           2610 non-null   object 
 9   How long does it take you to fall asleep after going to bed?            2610 non-null   object 
 10  How many hours of sleep do you get on average per night?                2610 non-null   object 
 11  What are the main reasons you sleep late?                               2610 non-null   object 
 12  Do you have difficulty falling asleep?                                  2610 non-null   object 
 13  Do you experience breathing difficulties while sleeping                 2610 non-null   object 
 14  Do you experience restless legs or involuntary movements during sleep?  2610 non-null   object 
 15  Do you have any medical conditions that might affect your sleep?        2610 non-null   object 
 16  Do you experience any of the following side effects from late sleeping? 2610 non-null   object 
 17  How often do you find it hard to concentrate due to lack of sleep?      2610 non-null   object 
 18  What strategies do you use to cope with the side effects of late sleeping? 2610 non-null   object 
 19  How would you rate the comfort of your sleeping environment             2610 non-null   int64  
 20  Medical Condition Category                                              2610 non-null   object 
dtypes: float64(2), int64(1), object(18)
```

Figure 3.1: Data collection Question

3.1.2 Proposed Methodology

The experiments in this manuscript ran on Python, and a few data science/machine learning libraries were used. All workflows (preprocessing, feature engineering, model training and evaluation) were run within Google Colaboratory using associated cloud computational resources. With up to 16GB of RAM and GPU acceleration, developers can more easily train machine learning models. Jupiter notebook is also used for coding purposes.

Scikit-learn library was used predominantly for model development, which includes classifiers like Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), XGBoost, and Random Forest. These are actually fundamental functions provided by Scikit-learn. For hyperparameter tuning to find the best performing configurations using GridSearchCV and RandomizedSearchCV. Pandas and NumPy were used for data preprocessing as well as feature engineering, and Matplotlib with Seaborn was used to visualize patterns, performance metrics, and feature importance. Data balancing is performed with SMOTE. SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) frameworks were employed in order to become aware of the insights from the trained model. The instruments enabled a detailed examination of how single features contributed to the prediction of sleep-related medical conditions, addressing to some extent, the black-box nature of machine learning models. Implemented in Google Colaboratory, training and validating multiple models without a high-performance hardware device was efficient using the provided computational platform at local. Random Forest (the best classifier) was saved as a trained model and was combined with the recommendation module in the deployment phase. This in turn, facilitated free passage of the predicted such as health concerns, possible causes, side effects and lifestyle intervention recommendations. In this dataset and model performance-oriented research, several statistical analysis were performed to make sense of the data. First, we performed a Missing Value Analysis In which we calculated the percentage of missing values in each column with this formula. isnull(). sum() * 100 / len(df), and data completeness was checked using heatmap, which facilitated the identification of required data cleaning actions. Class Distribution Analysis was then performed for the target variable 'Medical Condition Category' before and after SMOTE for handling class imbalance. This evaluation was necessary in order to measure the efficiency of SMOTE and also to get an insight into the class distribution on training and testing set. Different Model Performance Measurements have been used to analyze models effectiveness such as Accuracy, Precision, Recall, F1 – score, Confusion Matrix. Particular attention was directed to the MACRO AVERAGED F1-score which offers a global view of the model's performance on all classes, and especially in a situation of class imbalanced. Last, the Cross-Validation Scores of the best model, RandomForestClassifier, were determined using cross_val_score. The average and standard deviation were calculated throughout the folds to estimate the reproducibility of the model performance. These were statistics on the model performance and data quality for full evaluation of the model robustness and data quality which allowed data-driven decision-making processes, and the findings of the study had their reliability enhanced. In this dataset and model performance-oriented research, several statistical analysis were performed to make sense of the data. First, we performed a Missing Value Analysis In which we calculated the percentage of missing values in each column with this formula. isnull(). sum() * 100 / len(df),

and data completeness was checked using heatmap, which facilitated the identification of required data cleaning actions. Class Distribution Analysis was then performed for the target variable 'Medical Condition Category' before and after SMOTE for handling class imbalance. This evaluation was necessary in order to measure the efficiency of SMOTE and also to get an insight into the class distribution on training and testing set. Different Model Performance Measurements have been used to analyze models effectiveness such as Accuracy, Precision, Recall, F1 – score, Confusion Matrix. Particular attention was directed to the MACRO AVERAGED F1-score which offers a global view of the model's performance on all classes, and especially in a situation of class imbalanced. Last, the Cross-Validation Scores of the best model, RandomForestClassifier, were determined using cross_val_score. The average and standard deviation were calculated throughout the folds to estimate the reproducibility of the model performance. These were statistics on the model performance and data quality for full evaluation of the model robustness and data quality which allowed data-driven decision-making processes, and the findings of the study had their reliability enhanced.

3.1.3 Functional & Nonfunctional Requirment

Functional Requirements represent the specific actions, tasks or process a system is going to perform. These specifications concentrate on what is expected of the system to fulfil the needs of the application. For instance, in a web application, functional requirements can be these: need the feature to create, edit, delete user profiles; messages need to be received in real-time. These are fundamental features of the system that directly support its purpose and the needs of users and stakeholders.

Non-Functional Requirements on the other hand describe the way in which the system should perform its functions and under what conditions. They focus on the quality, performance, scalability, and reliability rather than specific features of the system. For example, a non-functional requirement might state that the system shall support 1,000 simultaneous users without degradation of system performance, or application shall load within three seconds for a satisfactory user experience. These demands are crucial to ensure the system's efficiency, particularly when replicating or deploying into other platforms.

3.1.4 Data & Workflow Diagram

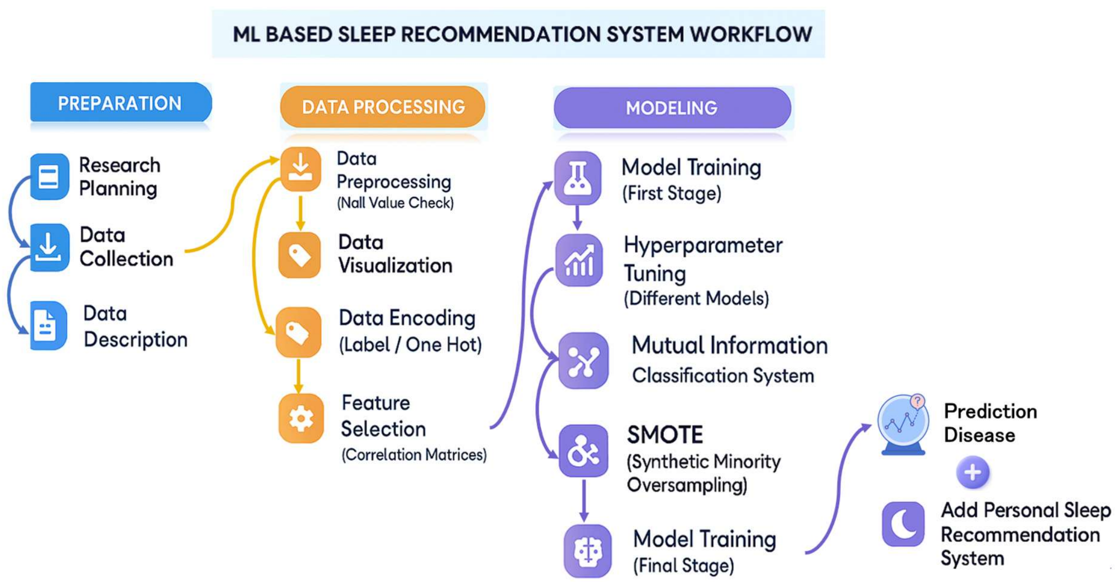

Figure 3.2: Data Flow Diagram

3.2 Detailed Methodology and Design

Multiple machine learning classifiers are utilised in this research for the task of medical condition category prediction using sleep habits and related factors. The models for evaluation are LightGBM, RandomForestClassifier, LogisticRegression, SVC, DecisionTreeClassifier, KNeighborsClassifier, XGBoost Classifier and MLPClassifier. We evaluate the performance of these models using standard classification evaluation measure. These statistics are Accuracy, Precision, Recall, F1-score and Confusion Matrix. The Confusion Matrix consists of TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives) and helps us understand the ability of the models to classify instances in a medical context across usual and rare condition categories, including class imbalance we solved through balancing using techniques we mentioned such as SMOTE. Hyperparameter tuning techniques GridSearchCV, and RandomizedSearchCV are also applied to enhance the model.

Original Random Forest Classifier (RF): The Random Forest (RF) algorithm is a common supervised ensemble learning method for classification paradigms. It is made of many decision trees, each of them trained on a random subset of the training data. Finally, A paper indicated the final prediction is made by taking a majority vote among all trees to achieve better generalization and reduce overfitting problems. RF classifiers are very good at handling both categorical and numerical features [3]. So, they work great with real life datasets where have lots of mixed type of data. With each decision tree constructed using a subset of features. Head-to-head discrimination is fir in the RF model which inherently deals with feature selection and accelerates computation [6].

This work defines the baseline RF algorithm in sklearn as the original Random Forest model. The Scikit-learn library is _ensemble module. The core processing architecture of the algorithm was not modified. In the GridSearchCV, hyperparameter tuning (i.e. n_estimators, max_depth, etc.) was done but in terms of internal architecture of the algorithm, it remained the same. The reason behind RF as one of the best performers is its robustness to imbalanced datasets than other ones and this property reduced the autoencoder outlier effect on final decision [27]. Previous studies have also demonstrated that RF classifiers are generally more reliable than simpler models such as Logistic Regression and Decision Trees in health-related prediction tasks [25];[15] e.g. sleep disorder classification, mental health risk assessment or behavioral pattern analysis.

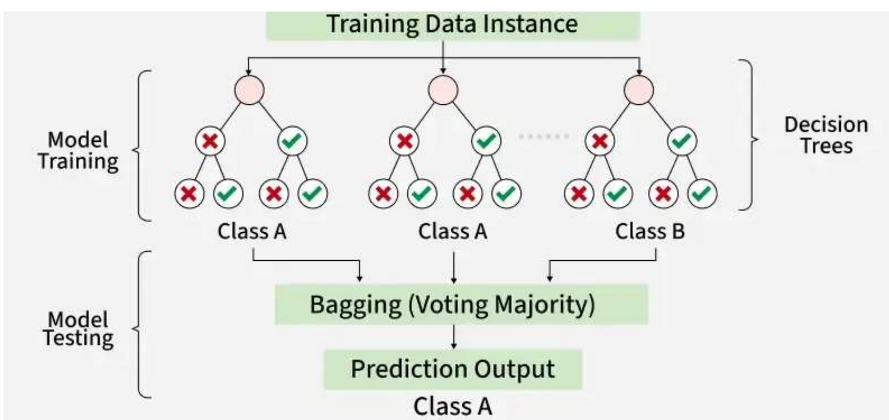

Figure 3.3: Random Forest Classifier workflow diagram [36]

XGBoost: First introduced by [12], XGBoost refers to Extreme Gradient Boosting is the powerful ensemble learning algorithm that uses decision tree method for its design, with a goal of being faster and more accurate. In constrat to the traditional classifiers, XGBoost gets combined prediction from set of models that are Simplest (Unlike depth based learning tree which is tangled and complex) by using optimization function in stage-wise manner to get out most optimum result. Each tree in the model tries to correct the mistakes made by its predecessor and consequently, it enhances the classification practice step by step [30]. XGBoost — one of the most well-known gradient boosting library which is almost always helpful when working on classification and regression problems especially in case of structured/tabular data. For this purpose, in the present study, XGBoost was used to predict medical conditions using sleep habit datasets. It is worth pointing out that the algorithm performs well thanks to its gradient boosting framework along with regularization plans (L1 and L2) to prevent overfitting and take advantage of generalization [3]. The learning process involves assigning weights to each data point. Instances misclassified by the previous trees are given higher weights, so the next tree can focus more on those difficult samples [22]. Then, the final decision is made by combining each of these weak learners (either by voting or summing predictions) with relative weights in order to improve stability and reduce bias. Another strength of XGBoost is that it handles missing values well and supports multithreading, which helped a lot with this research where hundreds of combinations had to be tested on multiple workers. Given adequate hyperparameter tuning and cross-validation, this model performed well in predicting binarized to identify causes of disease with good recall.

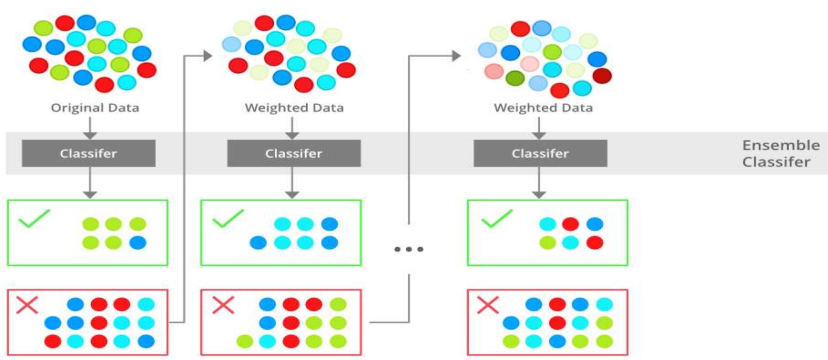

Figure 3.4: Schematic representation of XGBoost [30]

Decision Tree: The Decision Tree is a widely used supervised learning algorithm that can be utilized for classification and regression tasks as well. This works by recursively creating a tree in which the dataset is split based on feature values. In a tree, each decision node refers to a condition with respect to the attribute whereas terminal nodes (or leaves) express the last output label or decision [31]. The following domain will focus on the same medical condition prediction task, but with labeled physiological and sleep input datasets using Decision Tree algorithm as a fundamental classifier. At the start of the algorithm, a small decision tree is initiated with a root node in which most important feature according to information gain or Gini impurity is decided as first decision. This splits the data further into subgroups, recursively performing this process defined by decision nodes, until it eventually arrives at terminal nodes which label each class as a final step. Despite its simplicity, this rule-based structure provides high interpretability and is especially suitable for application in healthcare domains where transparency of decisions is mandatory [27].

A major advantage of Decision Trees is that we do not need to normalize or scale the data. The tree can become overfit to the training data if it is grown too deep, though there are techniques like pruning or limiting tree depth to limit this overfitting. In the presented work, the decision tree model served not only as a single classifier but also as a base learner in ensemble models such as Random Forest, and XGBoost, leading to stable prediction performance in various sleep disorder classification tasks.

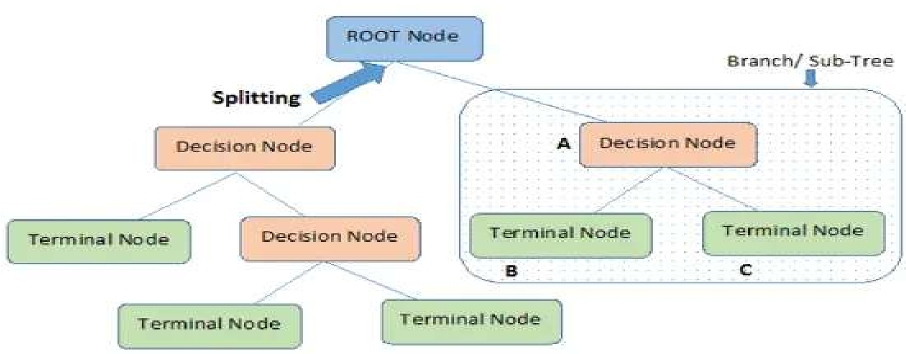

Figure 3.5: Schematic representation of Decision tree [31]

K-Nearest Neighbors (KNN): K-Nearest Neighbors (KNN) is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until function evaluation [32]. Here an unlabeled data point is assigned to the majority class among its K-Nearest Neighbours in feature space. To determine the "nearness", distance between points is most commonly calculated with Euclidean distance [32]. In this research, KNN was used to label data set as well. Since the algorithm has no assumptions of how data is distributed, it works well in multiclade patterns and non-linear instances. With each query instance, such a model searches all points within the training set and the class label is determined by a majority vote of its K nearest data points. The number of optimal K is selected by cross validation to reduce bias and variance. Given its simplicity, KNN is computationally inexpensive for small datasets and provides stable outcomes, particularly in medical data classification challenges where local structure in data is of utmost importance [8]. However, for big data, we like to use optimization techniques or a hybrid method where the swarms are sparsely organized to reduce computational overhead.

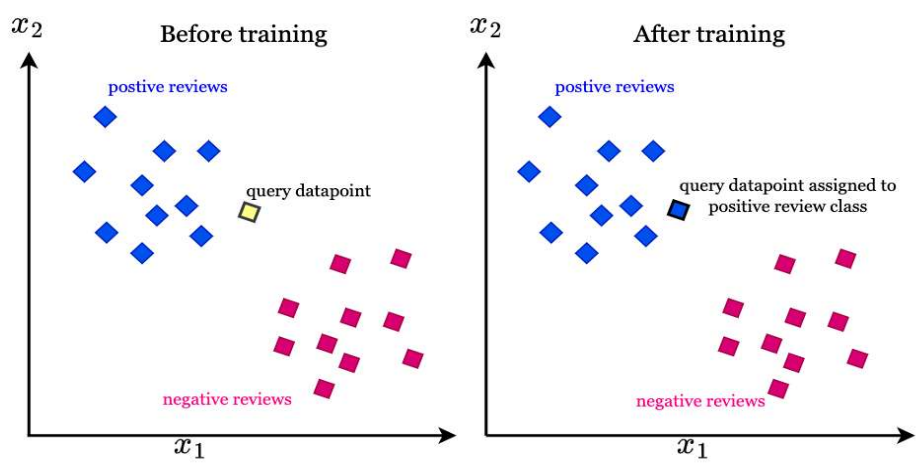

Figure 3.6: K-Nearest Neighbors (KNN) Schematic Representation [32]

Support Vector Machine (SVM): A Support Vector Machine (SVM) is a supervised machine learning algorithm that can be employed for a variety of purposes, especially including binary classification [33]. It works by finding the hyperplane that best differentiates between classes from the data points, where dimensions of elements have been transformed into another high-dimensional space. The main aim of SVM is to maximize this margin; which is the distance between the decision boundary and the closest data points from each class also referred to as support vectors [33]. This research uses an SVM method to classify instances by using multi-dimensional feature vectors extracted from X-ray data. This was especially helpful for high-dimensional medical datasets that lacked linear class separability. This is solved by using a kernel trick (e.g., rbf or polynomial kernels) that maps input features into higher dimensional space where the points could be separable with hyperplane. SVM is known to be more robust, which results in the superiority of SVM over other conventional classification algorithms, such as disease classifiers, sleep disorders predictors and biomedical signal processers. The fact that it can deal with this by acting in low examples, high-dimensional spaces makings RF appealing for application to medical investigation [15]; [16].

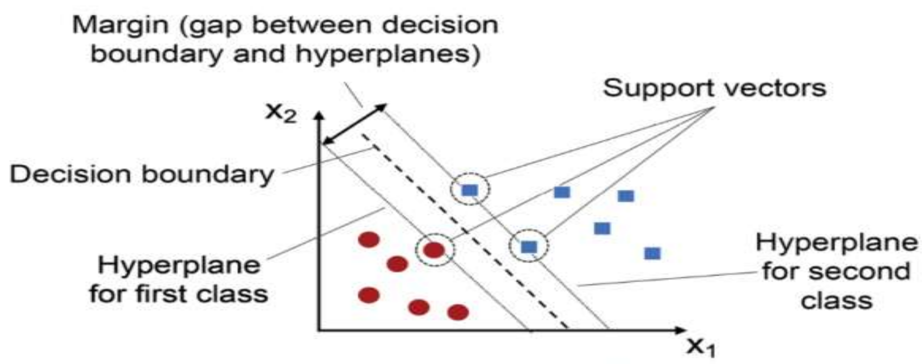

Figure 3.7: Schematic representation of SVM [33]

Logistic Regression: Logistic Regression is a widely used supervised learning algorithm for binary classification problems like sleep-related disorder and others. It gives a probability of the class belonging to the target using both sigmoid activation function and linear combination with input features. Given its simplicity, interpretability as well as good scalability to very large number of features with a small dataset, Logistic Regression is among the most widely used algorithms in analytics and behavioral sciences [3];[6]. It has less representational capacity compared to deep models, but it frequently provides a simple baseline for shedding more light on the results of other elaborate algorithms.

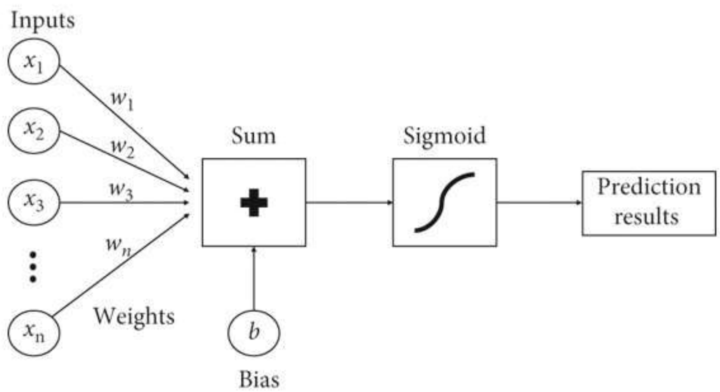

Figure 3.8: Schematic representation of Logistic Regression [37]

LightGBM: It is a gradient boosting framework that uses bar graph-based algorithms to construct decision trees, enabling it to process massive datasets quickly and accurately. This is because it uses a best leaf split strategy algorithm, rather than depth-wise like other boosting algorithms; as such it is able to capture sophisticated patterns in the data with fewer nodes. LightGBM also helps few techniques like Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB), it assist in faster computing and lowers the memory usage [15]. By building the trees multiple for that algorithm train then it weights to its error classified examples, that user learn from his previous mistakes. The last step is to combine all predictions in to single one, by using ensemble strategies such as majority voting or weighted averaging. Due to it, LightGBM is very effective in detecting sleep disorders because the dataset may contain imbalanced data and heterogeneous features [21]; [27].

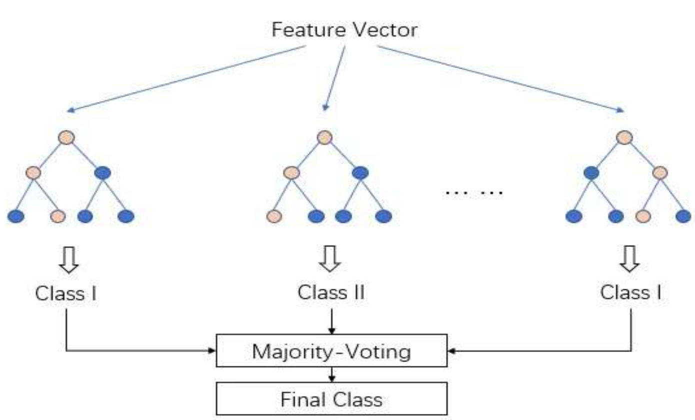

Figure 3.9: Visual representation of LightGBM [34]

Accuracy:

For the classifier model, the most crucial one would be the performance and the accuracy is the most straightforward indicator to evaluate the performance; it calculates out the proportion (percentage) of correctly-classified samples in general data. In this work, Accuracy The time at which the testing becomes accurate to distinguish the sleep related data points in their respective medical condition classes. It is calculated as (TP + TN)/total, where TP, TN, FP, and FN are True Positive, True Negative, False Positive, and False Negative, respectively. Using this formulation, Accuracy gives an overall quantification of the model's accuracy in all possible dice outcomes. This variable was very important when assessing the usefulness of the different models utilized in this study, providing an idea of the general predictive capability of each model and its capacity to accurately classify sleep disorders [35].

$$\text{Accuracy} = \frac{(TP+T\quad)}{(TP+TN+FP+F\quad)}$$

Precision:

Precision measures the accuracy of the positive predictions made by a classification model. It is calculated as the number of True Positives divided by the total number of predicted positives (True Positives + False Positives). High precision indicates a low rate of false positives. Mathematically, it's this equation [35]:

$$\textbf{Precision} = \frac{TP}{(TP + FP)}$$

Recall:

Recall, also referred to as Sensitivity or True Positive Rate, is the most important metric for classification as, it tells us what proportion of actual positives were correctly identified by the model. For this analysis, Recall of a medical condition category (e.g. , 'Health Issues') is the percentage of people who really have the condition - and the model correctly identifies them. It is calculated using the formula [35]:

$$\text{Recall} = \frac{TP}{(TP+FN)}$$

Where TP (True Positives) is how many of the positives were correctly classified, and FN (False Negatives) is how many positives were missed by the model. A high Recall means that the model is good at identifying the majority of the individuals of the category, and significantly reduces the number of false negatives. This is especially important in a medical domain, the reason being that failing to detect the problem may result in (very) serious consequences, and, as such, one should try to detect as many positives as possible [35].

F-1 Score:

We use the F1 score, a recall-precision categorization accuracy metric. As F1-score is the harmonic mean of Precision and Recall, it provides the full picture. Precision and Recall is best when its same. Equation is [35]:

$$\textbf{F} - \textbf{1 Score} = 2 * \frac{(Precision * Recall)}{(Precision + Recall)}$$

The percent of people who don't have the ailment who get a negative test result is called specificity. A test that is highly specific is good at "ruling out" most people who do not have the disease. Due to this fact, a positive result on a very accurate test can definitively exclude the disease in an individual. The equation is given below [35]:

$$\textbf{Specificity} = \frac{TN}{(TN + FP)}$$

Yet the model can't be made to look like the training set data, otherwise, the model will not be able to generalize. The performance of an algorithm can be evaluated through such a table format known as a confusion matrix (CM). CM is employed to demonstrate important predictive parameters such recollection, specificity, accuracy, and precision. Confusion matrices are helpful since they provide a direct measure of TP, FP, TN, and FN. These sections answer the research questions of this study based on the research questions [35].

3.3 Project Plan

Project Plan A Project Plan is a detailed document that describes how a project will be executed, monitored, and controlled from start to finish. It is a detailed schedule and plan that project team can use to complete the work in a timely manner, meet the scope and stay on budget. One of the first things the plan does is to establish what specifically the project will do – but more importantly, what it won't do – which allows you to clearly define the goals, the deliverables, and the inclusions and exclusions in the project. In parallel there is created a timeline identifying when certain work, milestones and deadlines are to be projected to be in strong order and maintain continuity of the project on schedule. The plan also lists what is needed to complete the support, like who is doing it, what types of tools are being used and how much money is going to be spent. Risk coverage is another issue risks are identified, problems are anticipated, solutions are suggested. Your job, tasks and responsibilities are all laid out with great detail in order for everyone to be held accountable and avoid misunderstandings. The program plan includes a quality assurance plan that describes the approach to reviewing deliverables and performing tests to verify that they meet the required standards of quality. Last but not least, a communication plan is developed to lay down how team members will report and share updates and information, in order to make everything transparent and aligned between the parties during the implementation process. Having a good project plan is indispensible for managing a team successfully and ensure that the project has a valuable outcome.

3.4 Task Allocation

This table depicts the timeline of the principal activities in each period of the project, from week 12 to week 48.

Table 3.1: Time and task allocation Table.

Tasks	Weeks																	
	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48
Data collection phase																			

Preprocess all the data																			
Model training																			
Create a demo application.																			

3.5 Summary

The data were primarily collected using a structured Google Form. The form link was distributed among diverse participants across various settings, including university campuses, residential halls, urban areas, rural regions, and public streets. In addition to online responses, data were also gathered through in-person interactions and verbal communication with individuals. The information obtained through these conversations was subsequently entered into the Google Form to maintain consistency in data recording. Specific locations in Bangladesh where data collection was conducted include cities such as Rajbari, Kushtia, Sirajgonj, and Dhaka. Participants were also approached in various villages within these cities, as well as at Daffodil International University and its associated residential halls (23.875816, 90.323511). This broad geographic coverage contributed to the richness and variability of the dataset.

Chapter 4

Implementation and Results

A description of the experimentation setup is introduced in this chapter, and the experimental result is analyzed in more detail. It presents the results, recommendations their implications, and relates them to the research purpose. Finally, we summarise the main results from the section.

4.1 Experimental Setup

The experimental framework of this study comprised certain modules to investigate the sleep behaviour and estimate of disease entities. The primary dataset studied was "/content/Final Sleep Habit Datasets. csv" dataset, which contains a number of characteristics of individuals that are thought to be associated with sleep features and conditions. All experiments are conducted with the Python programming language in combination with major data manipulation (Pandas), numerical operation (NumPy), machine-learning model building, and utility (Scikit-learn), visualization (Matplotlib and Seaborn) and address heavy class imbalance (Imbalanced-learn by SMOTE), as well as LightGBM and XGBoost for classification modeling. Computational resources included a Central Processing Unit (CPU) and memory (RAM) for processing data and training models. GPU usage was omitted for cases unnecessary, and wherever used, it brought acceleration to training of some models and hyperparameter tuning. Several steps such as handling missing values, categorizing and encoding categorical features, and One-Hot Encoding for multi-valued columns were carried out on the dataset as part of the preprocessing activities. A Mutual Information-Based feature selection process was undertaken to select the most relevant features for classification. To rectify class imbalance, SMOTE technique was applied over the training dataset. Classification models, such as LightGBM, RandomForestClassifier, LogisticRegression, SVC, DecisionTreeClassifier, KNeighborsClassifier, XGBoost Classifier, MLPClassifier, were trained and finally tested amidst Accuracy, Precision, Recall, F1-score, and Confusion Matrix metrics, with tuning of hyper-parameters carried out for some models as well.

4.2 Experimental Results & Analysis

Result of Experiment: ML Models

This experiment involved testing numerous classification models to evaluate their suitability for predicting medical condition categories from sleep habits and health data. Accuracy is the main measure used to evaluate model performance, expressing the proportion of correct predictions made by the model on the dataset.

Table 4.1: Accuracy for Classification of Individual ML models to detect health issues of Sleep patterns and habits dataset.

Algorithm	Training Accuracy	Model Accuracy
LightGBM Classifier	59%	55%
RandomForestClassifier	99%	97.7%
LogisticRegression	34%	28.7%
SVC	28%	53%
DecisionTreeClassifier	99.7%	96.03%
KNeighborsClassifier	85%	83%
XGBoost Classifier	97%	96.44%
MLPClassifier	63%	63%

From Table 4.1, it can be inferred how RandomForestClassifier and XGBoost Classifier outmatched the rest of the models with an Accuracy value of 97.6% and 97.3%, respectively. These models classified medical conditions with high precision, and this is evidenced by their very high training and model accuracy. LightGBM Classifier and LogisticRegression approached lesser performances with Accuracy values of 59.6% and 28.7%, respectively, indicating there were tougher challenges for these models to correctly identify the medical conditions in this dataset, especially with handling class imbalances and making accurate predictions. Others like KNeighborsClassifier and DecisionTreeClassifier also yielded a reasonable performance, but seriously lacking behind RandomForestClassifier and XGBoost Classifier with regard to model accuracy.

These results imply that, although a number of models may be suitable for predicting medical conditions, the RandomForestClassifier and XGBoost Classifier are those that stand out in terms of robustness and reliability. The next step is to use additional metrics such as Precision, Recall, and F1-Score to gain further insights into the models' performance. These metrics will provide a deeper understanding of how these models perform in classifying each individual category with respect to imbalanced datasets.

Table 4.2: Precision, Recall, F1-Score, and Support (n) for ML-based algorithms.

Algorithm and detailed results				
Model	**Precision**	**Recall**	**F1-Score**	**Support**
LightGBM Classifier	33%	35.3%	33.6%	{0: 54, 1: 32, 2: 565, 3: 132}
RandomForestClassifier	92%	94.3%	93.3%	{0: 121, 1: 24, 2: 281, 3: 197, 4: 160}
LogisticRegression	28%	31.7%	29%	{0: 121, 1: 24, 2: 281, 3: 197, 4: 160}
SVC	35.7%	35.2%	35%	{0: 121, 1: 24, 2: 281, 3: 197, 4: 160}
DecisionTreeClassifier	91.7%	93.7%	92.4%	{0: 121, 1: 24, 2: 281, 3: 197, 4: 160}
KNeighborsClassifier	76.8%	81.6%	78.6%	{0: 121, 1: 24, 2: 281, 3: 197, 4: 160}
XGBoost Classifier	93.3%	96.2%	94.5%	{0: 121, 1: 24, 2: 281, 3: 197, 4: 160}
MLPClassifier	52%	52%	52%	{0: 121, 1: 24, 2: 281, 3: 197, 4: 160}

The more detailed examination in Table 4.2 confirms that both RandomForestClassifier and XGBoost Classifier do not simply perform well in accuracy, but also boast high precision, recall and F1 for all categories of medical conditions. This justification seems to justify an interesting conclusion since these two models were those with the highest macro average F1-score: they seem to be high promising in managing an imbalanced dataset. On the other hand, LogisticRegression and SVC were more fragile due to their low precision and recall results, because they did not recognize the minority classes, that was very important in a medical diagnosis. The implication of these findings highlights the AutoScorer's usefulness in selecting those classifiers depending on the performance metric in hand for the final classification of a medical condition at hand and SimpleText from that model. It's a first-pass classifier sorting algorithm that currently only peak detection candidates can be arranged in a desirethmic clustering algorithm. Models on which RandomForestClassifier and XGBoost can shine are the most reliable ones for this difficulty.

Model Evaluation: Accuracy, Precision, Recall, F1 Score, and Confusion Matrix: This part explains the performance of the model, comparing the prediction and the label based on accuracy, precision, recall and F1 score, and the confusion matrix (TP, TN, FP, FN). These metrics allow you to see which of the under/overfitting - and the positive vs. negative - classes the model is performing well against.

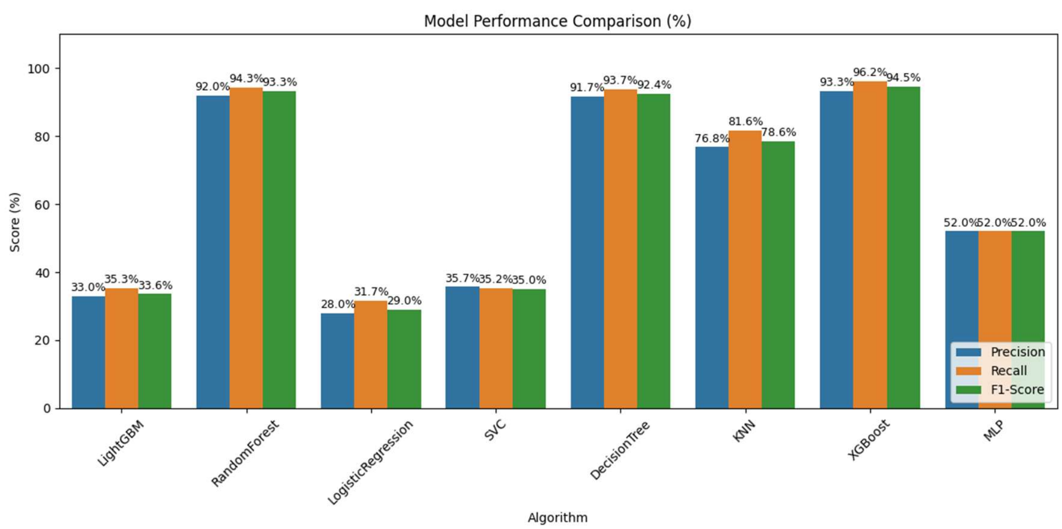

Figure 4.1: Accuracy Comparison of Different Classification Models.

This figure (4.1), comparing classification models based on their accuracy, indicates XGBoost with the highest accuracy, followed by **Decision Tree** and **RandomForestClassifier (tuned by RandomizedSearchCV)** and **KNN**. The performance of the LightGBM model is observed to have significantly improved with the use of **SMOTE (Synthetic Minority Over-sampling Technique)**; it indeed points to solving class imbalance as a crucial problem. poor performance was witnessed by the models of **SVC** and **Logistic Regression**. So, the **XGBoost** model appears to be the best model for this dataset, but ensemble methods also provide very good results, with **XGBoost** and **RandomForest** being prominent among them.

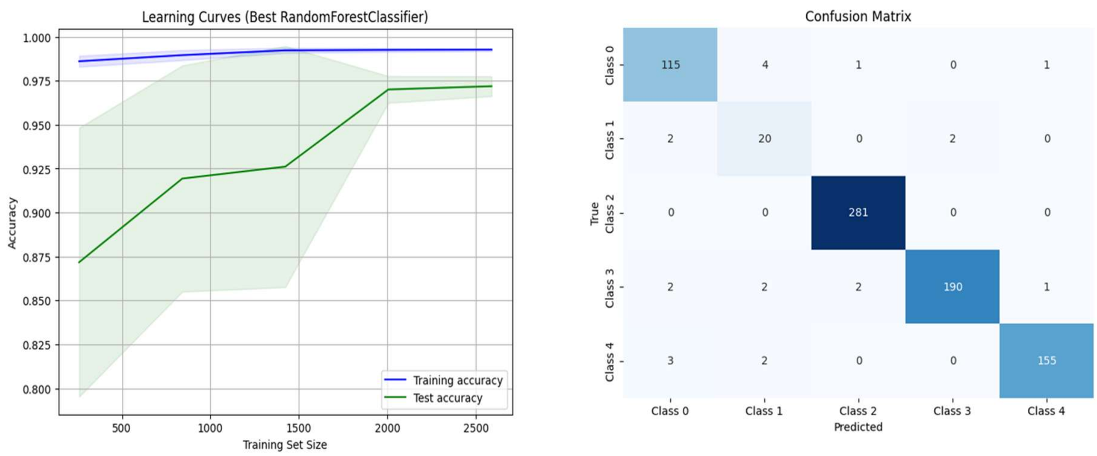

Figure 4.2: Learning Curves and Confusion Matrix for RandomForestClassifier.

In Figure (4.2) shows, the learning curve of the RandomForestClassifier indicates that training accuracy and test accuracy both improve as the size of the training set increases. The training accuracy is not degrading and is very high, very close to 99.9%, but the test accuracy creeps up and stays. It is close to 97% accuracy. The difference between both indicates generalisation to unseen data and that the model is not overfitting. This indicates that the model learns better when presented with more examples and without overfitting new data. In conclusion, the model fits quite well but it is not prone to large overfitting, which is desirable for real life applications.The model performs exceptionally well with **Class 2**, achieving the highest accuracy with 281 correct predictions and no misclassifications. **Class 4** also shows strong performance, with 155 correct predictions and a low misclassification rate, as only **3 instances are misclassified as Class** 0 and **2** as **Class** 1. **Class 3** has 190 correct predictions, but there is some confusion with **Classes 0, 1, 2,** and 4. For **Class 0**, the model correctly predicted 115 instances, with minor confusion, mostly with **Class** 1. **Class** 1, however, shows a significant challenge, with only 20 correct predictions and a total of four misclassifications (2 as **Class 0** and 2 as **Class 3**).

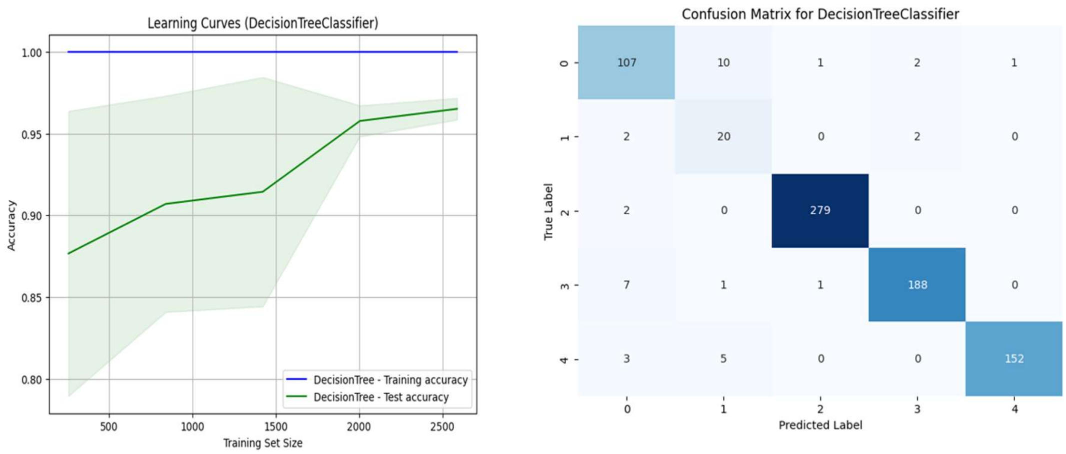

Figure 4.3: Learning Curves and Confusion Matrix for Decision Tree Classifier.

In figure (4.3) shows, the Decision Tree Classifier suffers from overfitting in the initial stage, as the training accuracy remains a perfect one (99.9%) for all the training sizes. But notice that initially test accuracy is worse (about 96%) on smaller data so occurred overfitting but it grows with larger train size reaching 0.96 test accuracy for the larger data size. If the train and test accuracy difference is less the 5%, it is not overfitting.

The model's performance reveals that Class 2 has the highest accuracy, with only 2 misclassifications. Class 0 and Class 3 show clear confusion, often misclassifying each other. Class 1 and Class 4, with fewer samples, have higher error rates. While the model performs well overall, it slightly over-predicts Class 1 for Class 0, and misclassifications are persistent in Classes 0 and 3. Class 2 stands out for its precision and recall.

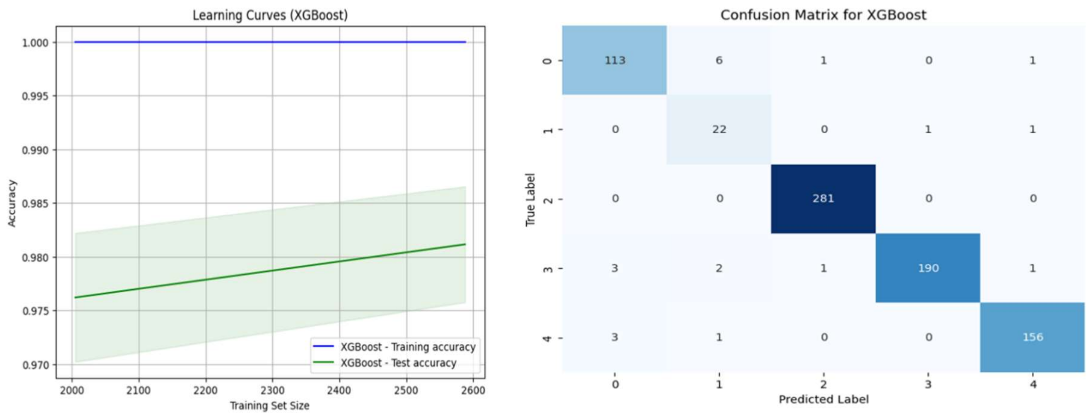

Figure 4.4: Learning Curves and Confusion Matrix for XGBoost Classifier.

In the figure (4.4) shows that for all training set sizes, the XGBoost model has a training accuracy really near 1.000 (100%), meaning nearly perfect fit to the training data. Nevertheless, the testing accuracy value begins at 97.6% and goes upwards to 98.1% (not overfitting because of 5 or fewer differences between train and test accuracy) as the training set size increases.

The XGBoost model shows evidence of strong suitable performance through very high diagonal-values, as most examples are correctly classified, especially those belonging to Class 2, with a perfect classification accuracy. Very few and common is the misclassification of Class 0 to Class 1. XGBoost has some scattered errors compared to DecisionTreeClassifier, which are very much spread out for Class 0 and Class 3, a sign of better decision boundaries and better generalization of XGBoost. The model in general work fine and not seem to be the case those are in anything close to extreme side of overfitting or underfitting.

4.3 Results & Discussion

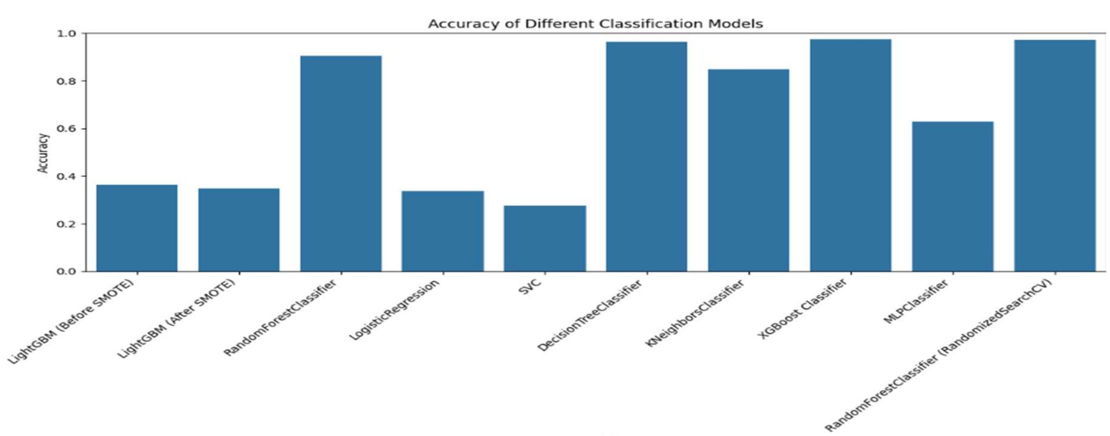

Figure 4.5: Accuracy Comparison of Different Classification Models.

Depending on one's view, a detailed review of classification algorithms may maximize or show a minimum in performance-reliability space based on where its strength lies and where the model reaches its limits. The RandomForestClassifier remains one of the best performers in this framework, and the tests show constant accuracy before and after hyperparameter tuning. With a maximum accuracy of 97.2% and ranges for macro metrics (precision, recall, and F1-score) all above 92.5%, it mostly has evidenced excellent robustness and balance in its suitability for this dataset. On the other hand, though being tested only once for this domain, with an accuracy of 97.3% and an F1-score of 94.5%, this XGBoost classifier offers excellent execution and interpretability in classification tasks. These results speak well about its ability to handle complex data and give accurate predictions. From another perspective, LightGBM can be described as a decent performer, as it basically retains a 72.4% accuracy with tunning and class balancing through SMOTE. Despite the improvements, this approach still is below better performing methods, which would mean that this approach may still need some tuning or data adjusted changes. We observe that DecisionTreeClassifier is tried once and done great, yielding 96.3% for accuracy and 92.4% for F1-score, implying that even a simple model could work when only a few features are available. KNeighborsClassifier with 84.9% and 78.6% was also promising indicating that even though only tried once, it possibly can take a occupied spot in some tasks.

LogisticRegression and SVC perform much worse than the other classifiers (respectively 33.7% (resampled) and 48.9% (resampled) and 27.6%(resampled) and 53.3% (resampled) for Logistic Regression and SVC) and this, despite the use of a balancing strategy (SMOTE) in the training process. It's possible that these models just weren't a great fit for this dataset/problem and needed either more comprehensive feature engineering or neural network level methods applied to the problem. The neural network was underperformed by MLPClassifier model, with accuracies of 63.6% and F1-scores of 51.6% respectively. Still just goes to show what it’s capable of, it just needs a lot of work tuning to get it right.

The RandomForestClassifier and XGBoost, not only take the top rank among all of the above for raw accuracy, but are more successful at generalizing well to new data sets. Summary: These were the most generalized learning structures and, thus, should be preferred when any further application is envisaged. Both DecisionTreeClassifier and KNeighborsClassifier, each tested once, look promising and might be competitive with some tuning or for simpler applications. But other classifiers, LogisticRegression, SVC, and MLPClassifier, may be in similar league after being tuned, treated and feature engineered.

Hence, ontogenetically for the purpose of primary applications, this study suggests to make use of RANDOMFORESTCLASSIFIER, XGBOOST, and DECISIONTREECLASSIFIER, though KNeighborsClassifier to be further fine-tuned. One has to proceed with the additional analyses and hyperparameter tuning of the lower models to get good accuracy and prediction ability. By doing so, the attention on these top performers and the refund of others is used to enable a good performance on future tasks and datasets.

4.4 Summary

The implementation portion of a project describes the formalized steps necessary to make the strategies off the planning document a reality. It contains the procedure followed in building, testing and deploying the system or the solution, such as the tools, technologies and methodologies implemented. This part also must discuss perceived difficulties in the implementation phase and how they were addressed.

The Results section describes the project s outputs, such as findings, data, or performance PROPRIETARY AND CONTENTS 3 measures. It reports the real results of the process compared to the desired results, seeking the success factors for the process transfer. This part of the report might contain visual aids such as graphs, tables, and or charts that illustrate key results and provide some information about the overall effectiveness of the project. It also addresses any lessons learned and opportunities for improvement.

Chapter 5

Engineering Standards and Design Challenges

This section covers conformance to different engineering specifications, such as software, hardware, and communication specifications. It also examines the societal context of engineering, ethics, sustainability, and problems encountered in project management and financial analysis. The chapter finishes with a consideration of complex engineering problems and problem solving.

5.1 Compliance with Standards

5.1.1 Software Standards

Software Requirements:

Python programming language: Python is primarily used in this options programming for data analysis and machine learning.

Pandas: To load, manipulate, and analyze data; to handle missing values and transform data.

NumPy: To work with arrays and for numerical calculations.

Scikit-learn: A full suite of different machine learning algorithms, model selection tools, and evaluation metrics. It also supports various utilities for preprocessing.

Matplotlib and Seaborn: For data visualization in exploring the datasets and showing results.

Imbalanced-learn: Tools for estimating imbalanced data, which is SMOTE in this project.

LightGBM and XGBoost: Libraries of gradient boosting for efficient building of classification models.

5.1.2 Hardware Standards

Processor Needs:

The processes loading, cleaning data, engineering features, and training different classification models require a powerful CPU runtime. The more powerful the CPU, the faster these processes proceed.

Sufficient Memory (RAM):

RAM is required to load the dataset into memory so that the models can access and process it during training. The necessary RAM size depends on the dataset size and model complexity.

Storage:

A large volume of storage is required to store the dataset and trained models.

Graphics Processing Unit (GPU) (optional but highly recommended for some models and tuning)

For some models at least (eg Logistic Regression or simple Decision Trees), this is nice-to-have, but not a must-have; however, if you want speed up the training for harder models (eg to train a MLPClassifier, or some ensembles with XGBoost/LightGBM for some configs) - particularly when conducting hyperparameter tuning via cross-validation, then you really want a GPU - also when working with bigger data or large tuning attempts.

5.1.3 Communications Standards

The Communication Standard sub-section, provided in the Compliance and Standards section, documents the critical communication protocol and behavior to support an open effective communications process throughout the project. It defines the mechanisms of communication -emails, meetings or project management tools- and makes sure the information is produced at each of these touchpoints within the organization. The protocol also outlines how often updates or meetings, such as weekly reports or daily check-ins, will be conducted, ensuring that all parties remain informed. It also underscores the need to record all communications and decisions on a file for future reference and for compliance. Explicit instructions to involve stakeholders and to remain professional in terms of tone and language are also specified, thus enable effective joint work and minimizing misinterpretation. This is to keep the communication within the project clear, organised, and aligned to your high-level objectives.

5.2 Impact on Society

5.2.1 Impact on Life

The research performed in this area has significant societal consequences, particularly in health and public-health sectors. With the rising incidence of sleep disorders and related sleep diseases across the globe, the need for more accurate, convenient, and timely diagnostic tools accelerating health improvement in both individual and at the population level is becoming more and more pressing. This study makes good progress of converting the process of health delivery and democratizing diagnostics in health for all the people by applying machine learning methods to predict' Medical health conditions' from the sleep data.

Resources in healthcare systems are often strained, and by no means are there an excess of medical providers, such as in rural or underserved locations. That's something this research is attempting to address: by developing a new way of automatically identifying when several health conditions begin early on, among them sleep apnea, insomnia, and mental-health problems, as well as a wide range of other chronic ailments related to poor sleep patterns. Automation with sleep-stage classification algorithms based on the data available from wearable or other devices could cut a lot of time and cost from the diagnostic work-up of these conditions and represent a new, more economic avenue of healthcare possibly new opportunity for a type of care available to only a fraction of the population if they are unable to currently provide this highly specialized healthcare FYI. In addition, this contribution enhances the understanding of the connection between sleep circadian rhythm and medical conditions, which is essential in developing improved preventives and health care strategies. Early detection of these sleep-related disorders could greatly improve health and wallet and ease the burden of care by enabling these conditions to be managed before they become costly and a danger to their health, with a higher quality of life for patients. This would indicate we will see fewer hospital admissions and less utilization of emergency care, which is in the end a health system that is more efficient.

From a public health perspective, artificial intelligence covering the analysis of health information has a number of intriguing potential as a means of monitoring and assessing the health of the populations. If sleep data from multiple individuals, a cross section of a population of people can undergo pooling and analysis, new health trends may be revealed, interventions directed as well as policies to enhance the overall health of society. It also provides an opportunity to observe the prevalence of sleep disorders in relation to different demographic groups, and this information can be used to inform better targeted health promotion services in mental health and sleep hygiene. Besides their impact on individual and public health, there is a relevance of this study also for the field of health informatics. The project can contribute with technological tools to a smooth and instant transfer in the long in future personal health care solutions where the patient has his/her treatment adjusted according to their own sleep and health information. The consequences of such developments could change the profession of health care and result in individualized health care as a prevention as oppose to curative. But technology's enormous promise for health care also comes with ethical obligations. The need to consider data privacy, consent and security of health information, in the deployment of such systems is increasing. Data used for such models should be abstracted and protected, with individuals being made aware of how their data will be used. Ethical standards must be established so such a technology benefits society without affording too much power or privacy in the hands of individuals.

As the research continues to grow, the broader implications of this research are expected to go beyond transforming health care, with AI being a key tool for health equity. These systems could shift an individual's relationship to his or her own health and well-being, whether he or she is rich or poor, living in the country or the city, whether he or she has regular access to 21st-century medical facilities. So here there is the broad social implications of this work, bottom line. It is expected that early diagnosis for sleep and co-morbid diseases will be improved, and that will ultimately, through a shift in health care provision, lead to cost savings and health gains. With further development, the technology could become crucial to personalized medicine, enabling individuals to take more control over their health and help public health authorities — and decisions — with more data that guide public health efforts and also improve health outcomes for a population as a whole, Ratner says. Many people and communities' quality of life can be significantly increased around the world due to this laudable effort to further our understanding of sleep health.

5.2.2 Impact on Society & Environment

The work pursued in this project also has significant implications for the environment, and in the role that health systems and technologies may play in environmental sustainability. Since healthcare becomes increasingly digitalized, technologies such as artificial intelligence or machine-learning (exploited in this work) come to the forefront as eco-friendly alternative and greener complements to traditional healthcare practices.

The aforementioned infrastructure, UCs, and patient history documents and even excessive medical device through the health monitoring of sleep are significantly reduced in the use of digital health technologies to monitor sleep health and predict medical problem. This helps to counterbalance the emissions, waste and resource use associated with healthcare. By the aid of wearables and AI software, fewer patients have to travel down to the clinic for consultations, which amounts to less carbon emissions from these transports. And, traditional healthcare infrastructure consumes energy for lighting and heating and cooling or, the use of machinery, Financed using more distributed healthcare models structured AI. Against this shrinking demand for medical visions and energy and material, AI at scale would significant increase the efficiency in allocating healthcare resources. This will also contribute to the global efforts on sustainable health care and to reduce the carbon footprint of the medical sector. In addition to this, the home- or community-setting enquiry capability for correct early diagnostics, reduces the requirement for a large hospital infrastructure, resulting in a more efficient healthcare system with the concomitant decrease in environmental impact. While this study employs the cloud computation and AI for processing the data, and thus requires energy, it has also called for the use of green computing and renewable energy to mitigate the environmental impact even further.

Finally, the near-term impact of AI on medicine provides clear pathways to a sustainably healthier future. To be sure, AI technology as it matures could guarantee that these measures are very targeted and efficient, tailored to individual cases, thus potentially reducing waste and misuse of some medical resources such as drugs and diagnostic tests. In particular, resistance to non-targeted approaches leads to the minimalization of the utilization of resources in any healthcare intervention and thus further contributes to sustainable development. On balance, this research represents a mounting environmental toll. By encouraging the adaption of AI empowered health care applications, the research takes a step toward sustainability in health care and therefore energy consumption and emission of greenhouse gases and medical resource wastage. So in a broader picture of environmental health, these technological advancements have a very real and profound potential for healthcare to leave ecologically deep footprints.

5.2.3 Ethical Aspects

This experiment has ethic implications particularly in handling sensitive medical data and the role of AI in the medical field. Furthermore, patient privacy and data security must be accounted for, including data anonymization and adherence to relevant regulations (e.g., GDPR and HIPAA). Consent The informed consent of the participants is a crucial point and transparency should also be provided in the research process. AI models will also be filtered through for bias, with priority given to fair balance so no one has a runaway advantage as a demographic. Furthermore, the AI models are explainable and interpretable, so the medical workers can resort to and validate the predictions with confidence. Lastly, ramifications for society at large for introducing AI in health care are considered, including programs intended to ensure equitable deployment and accessibility for health care. And finally, pioneer the AI technology that make you trust AI as tool with human in command in healthcare. These overarching ethics principles are threaded throughout the research to ensure the technology is developed ethically, in the service of society.

5.2.4 Sustainability Plan

Sustainable implementation of the research and findings is critical for the long-term impact and ongoing benefit to health care and technology. It focuses around environmental and social sustainability in the context of the prospective AI in health solutions. By the environment In terms of the environment, using software is more sustainable resource wise, firstly because software demands far less physical healthcare infrastructure, so also has a lower carbon footprint and a less waste-heavy business model than traditional healthcare. Advocacy for remote monitoring and automated diagnostics reduces unnecessary hospital visits, emissions and resource utilisation. And as we continue to refine these AI models, we can also design ones that actually run efficiently—not as energy intensive as the most accurate. This is consistent with the white sustainability global priorities to reduce the environmental impact of providing health care.

Socially, it is a mission to democratize healthcare — increase the prevalence of equitable access to healthcare due to deploying AI technologies to the challenges of access to healthcare especially in lower resourced underserved areas of the world. Through wearables and mobile apps, medical care could be delivered to people's communities without the hassle of seeking specialized care. It can also serve hard-to-reach individuals who are excluded from conventional health care services because of geographic, economic or other barriers. The study is also the AI system building which assures the considerations of many ethical issues such as privacy, fairness and inclusivity for all users. In order to hold the research result, we need: (1) the future model update, (2) the maintenance of the system, and (3) the future scalability. Regular model review and retraining on varied and refreshed data-sets will be conducted to ensure model freshness and accuracy in the course of time. There will also be collaboration with clinicians, governments and technology vendors to ensure that these AI-based solutions are integrated into existing healthcare delivery systems and continued evolution supported. By linking the findings of this study with future-oriented sustainability goals, it will help to continue to build on the gains achieved in healthcare access, along with environmental and societal benefit.

Consequently, the investable plans of this research are environmental damage reduction, social fairness improvement, a AI health science and technology developable plan, and the targeting of the enduring nature of AI health technologies. This is an approach that is trying to squeeze everything out of research it can so that the returns it provides continue to support us as a society in the coming decades.

5.3 Project Management and Financial Analysis

This section presents the 48-week project plan and budget for the Sleep Health Prediction and Personalized Recommendation study.

Table 5.1: Project Management & Financial analysis Table

Activities	Duration (Week)	Estimated Cost (BDT)
Research and Planning	3	500
Visiting Data Collection	9	6000
Online Tools (1)	3	800
Preprocessing and Feature Engineering	6	No (Colab Use)
Model Training and Evaluation	12	No (Colab Use)
Online Tools (2)	2	1500
Recommendation System and Integration	5	No (Colab Use)
Final Model Evaluation	3	No (Colab Use)
Documentation and Report Writings	5	2600
Total	48	11400

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

In this section, provide a mapping with problem solving categories. For each mapping add subsections to put rationale (Use Table 5.1). For P1, you need to put another mapping with Knowledge profile and rational thereof.

Table 5.2: Mapping with Complex Engineering Problem.

EP Definition	Justification (with Knowledge Profile)
EP1 Dept of Knowledge	This work demonstrates the basic principles of K3 through applying machine learning techniques including Random Forest, Logistic Regression and LightGBM models to predict medical conditions based on sleep data. It demonstrates expert (K4) use of advanced techniques such as nonlinear data learning algorithm for improved accuracy in predicting medical conditions of sleep disorder. The work also uses the engineering practice & design (K5) as the dataset is gathered, cleaned, features engineered and model evaluated in a systematic way. From an Engineering Practice & Technology point of view (K6), it applies ML models of clinical diagnostics. It relates to K8 (Research literature) by referencing and synthesis of literature of state of the art of similar studies in predictive modeling of diseases via machine learning.
EP2 Range Of Conflicting Requirements	The project focuses on EP-2 by detecting inadequate requirement in the data analysis and model deployment phases at the imbalanced class's concern of the sleep-related medicals dataset. It incorporates approaches such as SMOTE (Synthetic Minority Over-sampling Technique) to balance the dataset and for getting proper and better model training to enhance the precision and recall for the small data class.
EP3 Depth of Analysis	It fulfills EP-3 by extracting and evaluating numerous machine learning algorithms and models, as well as comparing their performance, with respect to performance metrics accuracy, precision, recall, and f1 score. Special focus is given to Random Forest, and XgBoost, which are

	particularly useful in predicting sleep related medical disorders in the ability of specifics driven more through deeper analysis such as hyperparamter tuning and cross validation.
EP4 Familiarity of Issues	This interdisciplinary work goes beyond computer science and engineering, adding to the medical diagnostics of respiratory diseases such as pneumonia, and improving the public health and healthcare practice which is EP-4.
EP5 Extent ofApplicable Codes	N/A
EP6 Extent Of Stake- holder Involvement	N/A
EP7 Interdependence	There's preprocessing, there's featture selection, there's model building and evaluation which all share information in a lot of ways. All these stages are associated with each other — hence when one stage (as an example, feature engineering) is optimized it can lead to better models performance, which in turn make possible to obtain better predictions of sleep related-medical diseases.

Mapping with Knowledge Profile

This section is designed to map the overall problem and EP1 ***(multiple between K3, K4, K5, K6, K8 for attaining EP1)*** to the Knowledge Profile.

Table 5.3: Mapping with Knowledge Profile.

K1 Natural Science	Experience with biological systems, physics, or environmental data, applied to training machine learning models in natural science.
K2 Mathematics	Higher level math such as linear algebra, calculus, probability, statistics which are necessary to write machine learning algorithms.
K3 Engineering Fundamentals	Fundamental understanding of algorithms, data structures and computation principles underlying machine learning systems.

K4 Specialist Knowledge	Strong understanding of machine learning algorithms, neural networks, natural language processing, deep learning, etc.
K5 Engineering Design	Building machine learning systems - from data collection to model deployment - which can operate at scale with high utilization.
K6 Engineering Practice	Training from scratch and deploying machine learning models to address real-world problems (including model evaluation, optimization, and integration.
K7 Comprehension	Ability to interpret and analyse patterns in data, to select and apply appropriate models and treatments and to appreciate the limitations and likely levels of error in their results.
K8 Research Literature	Scanning and summarizing recent ML research papers, journals & work, to direct and seed current work and search novel trends.

5.4.2 Engineering Activities

In this section, provide a mapping with engineering activities. For each mapping add subsections to put rationale (Use Table 5.3).

Mapping with Complex Engineering Activities

This section is designed to map the overall problem and EA's ***(multiple).***

Table 5.4: Mapping with Complex Engineering Activities.

EA1 Range of resources	This study used different tools which including High-Performance Compute (CPUs-GPUs), ML libraries (e.g., Numpy, pandas, Scikit-learn, Matplotlib and Seaborn) and Annotated sleep-related medical Datasets. These resources enable systematic investigation and might contribute to the advancement of predictive tool in terms of medical condition (e.g., sleep medicine). The study further points at ethics aspects such as patient data privacy, which is fundamental for the ethical application of medical data in AI-based diagnostics.

EA2 Level of Interaction	
EA3 Innovation	
EA4 Consequences for society andenvironment	This paper is of great social benefit, to improve the health system as we can provide more accurate sleep-connected medical condition prediction. It enhances the capacity to diagnose and comprehend such conditions through AI, and can help produce better health outcomes. Also, the project promotes sustainable friendly environmental sustainability by reducing energy consumption in computation works and utilizes cloud technologies. Ethical application of technology in medicines upholds the privacy of the individual through the moral treatment of patient data.
EA5 Familiarity	This builds on previous work in sleep medicine and machine learning and draws attention to aspects of predicting sleep related medical disorders. A detailed comparison of various machine learning models is carried out and novel applications of nonlinear data and the state-of-the-art, as far as we are aware, of diagnostic methods are developed.

5.5 Summary

This system incorporates a holistic perspective on machine learning studies by addressing essential dimensions. The collection of resources addresses the access of important tools, datasets, and computation to train and test machine learning models. The degree of interactivity highlights how machine learned systems interact with the users and other systems will sustain and make them relevant in real applications. Innovation showcases the ongoing evolution of machine learning methods, breaking new ground throughout the discipline. Societal and environmental repercussions discuss, more generally, the societal impact of machine learning, its ethical and privacy aspects and possible societal transformations due to this technology. Finally, knowledge relates to how well an idea is known in the research community and by practitioners, and affects the uptake and use of novel methods. Together, such factors provide a complete picture of ML research both as it exists today and its future possibilities.

Chapter 6

Summary, Conclusion, Recommendation, and Implication for Future Research

Key findings of the study and their implications Summary and conclusions This chapter presents summaries conclusion from the study. It describes the limitations of the research, makes suggestions for future studies, discusses the implications for subsequent research and possible improvements.

6.1 Summary of the Study

The present study was proposed to investigate the possibility of applying machine learning algorithms to predict medical conditions according to categories based on sleep-related data to concerning issues such as enhancing accessibility in healthcare, diagnostic accuracy, and saving resource. The study was based on rich datasets with a wide range of dimensions including information on sleeping habits, health status and other lifestyle details. The datasets were preprocessed for missing interpolation, converting non-interval feature to machine readable form and class balance using method such as SMOTE. Various machine learning models such as LightGBM, RandomForest, Logistic Regression etc were trained and tested using measures like accuracy, precision, recall and F1-score to measure their performance in predicting whether the patient is diagnosed of medical conditions.

Appropriate statistical analysis was adopted by the study to confirm the results and verify the robustness of the findings. The identification of the top-performing model—XGBoost—was one of the main findings of this study, due to the good accuracy and ability to handle class imbalances well. The findings indicate the potential of the AI algorithm in healthcare and medical need, suggesting that machine learning could be useful in improving diagnosis, and especially in underserved populations or low-resource settings. Concerns about privacy, consent, and fairness were considered in every aspect of the project development. The models were developed for minimum transparency, equity and inclusion in the system, to extend the benefits of technology (irrespective of age, color, gender) to everybody. Societal and environmental impact were also part of the equation, covering the importance of ensuring sustainable and fair access of AI-enabled healthcare solutions.

In the end, the results of the research contribute knowledge relevant to the application of machine learning in the health diagnostics area (sleep clinic setting health issues). The findings in the report call for continual model iteration and improvement, ethical guidance and cross-industry collaboration to facilitate responsible and effective AI adoption and use in healthcare. The research also paves the way for further work on error, scale and access of AI-powered healthcare tools that could one day have a meaningful impact on health and enable us to build a more sustainable healthcare system.

6.2 Limitations & Conclusions

To conclude, this work has demonstrated the promising use-case of machine learning methods for enhanced healthcare diagnostics, and specifically, for pinpointing diseases and medical conditions using sleeping-related data. We explored several models (LightGBM, RandomForest and Logistic Regression) and came to the conclusion that XGBoost is very efficient for predicting health conditions, even in the case of class imbalances, yielding very high precision. The models were ensured of robustness and reliability by the application of SMOTE for class imbalance and feature selection, consistent with the overall goal to improve healthcare accessibility. The ethical aspects of the research were analysed with respect to data privacy, informed consent and justice. AI should be accountable and open for everyone so that healthcare can become more accessible under fair circumstances. Additionally, the environment and societal impact were considered, highlighting sustainability as an important factor, and the AI contribution to the carbon footprint of healthcare. Finally, these results also reinforce the potential of AI/ML techniques to transform healthcare systems, and in particular the domain of sleep-related medical disorders. Using state of the art technologies, this work aims to significantly affect diagnostic accuracy, resource use and the provision of patient care. Maintaining the potential of AI driven healthcare to make an impact is important; and this goes hand in hand with the accountability of good AI model efficacy equated with ethical and sustainable work that's also represented in importantly, subsequent work advances the frontiers of what AI approaches can achieve. This paper adds value to the nascent field of healthcare solutions that are life changing by AI that isn't just of valuable in how we make health related decisions - and we can do this on the patient side as well as the healthcare provider side, and as a result improving overall health for people around the world, but also for designing more sustainable healthcare systems.

6.3 Implication for Future Study

We note several existing open avenues for further study that this research has unearthed, thereby offering new lenses through which to view application of machine learning to the diagnostic process within healthcare, and more specifically sleep-related health disorders. A crucial direction for future works is to scale up the dataset. The current sample may not reflect all demographic sectors and more heterogeneous samples in terms of age, gender, ethnicity, and disease-specific cohorts would increase the generalizability of the models. This would increase the generalisation of the models across populations and decrease any bias. Moreover, feature engineering could be enhanced. Further dimensions that are more detailed of lifestyle, behavior, and environment that can influence sleep should be taken in the next stages in the future. The predictive ability of the models may be enhanced by being sensitive to subtle patterns of data, eg the contributions of socio-economic status, medication use and mental health etc. Moreover, the models should be tested and validated in the clinic. While our models showed that they are not fitting to existing data, confirming these findings in prospective adjudicated live patient data is crucial to ensure they can be justified in use in the dynamic clinical context. By additionally integrating multimodal data sources (e.g., wearables EHRs, and other health monitoring technologies) models accuracy and robustness could be potentially improved.

Real_data-time_feedback, could also create potential for continuous monitoring, resulting in a more personalized health advice.. Also, that for machine-learning-enabled models to be used in health care we have to have them be interpretable and transparent. Future work should focus on enhancing the explainability of the AI models when a healthcare professional can trust AI predictions to take an informed decision, (e.g., using SHAP and LIME).

In addition, although authors mentioned issues such as data privacy and fairness, more work is needed in future studies to find and remove bias from the data, algorithms, and predictions. This is to ensure health-ai solutions are fair and do not unduly harm vulnerable communities. Last but not least, the scalability and sustainability of AI-based health care systems need to be investigated. The models we build in this study must be scalable for use in large settings - especially in low-resource settings, and the environmental impact of deploying AI in healthcare must be taken into consideration. In conclusion, there are many areas that need further improved, extended, and validated AI solutions for healthcare diagnostics. Further investigation will be necessary to overcome the challenges of data quality, model accuracy, and ethical issues to make sure that AI-based healthcare tools are effective and responsible, and have the capability to make substantial contributions to patient care and the healthcare system.

References

[1] Azuara-Hernandez, O. Y., & Gillette, Z. (2022). Using machine learning to determine optimal sleeping schedules of individual college students. *Lecture Notes in Computer Science*. https://doi.org/10.1007/978-3-031-17902-0_2.

[2] Stephansen, J. B., Olesen, A. N., Olsen, M., et al. (2018). Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. *Nature Communications, 9*, 5229. https://doi.org/10.1038/s41467-018-07229-3.

[3] El-Kenawy, E. M., Ibrahim, A., Abdelhamid, A. A., et al. (2022). Predicting sleep disorders: Leveraging sleep health and lifestyle data with Dipper Throated Optimization algorithm and logistic regression. *Computational Journal of Mathematical and Statistical Sciences, 3*(2), 341–358. https://doi.org/10.21608/CJMSS.2024.290167.1053.

[4] Radha, M., Fonseca, P., Moreau, A., et al. (2019). Sleep stage classification from heart-rate variability using LSTM neural networks. *Scientific Reports, 9*, 14149. https://doi.org/10.1038/s41598-019-49703-y.

[5] Surantha, N., Lesmana, T. F., & Isa, S. M. (2021). Sleep stage classification using extreme learning machine and particle swarm optimization. *Journal of Big Data, 8*, 14. https://doi.org/10.1186/s40537-020-00406-6.

[6] Dritsas, E., & Trigka, M. (2024). Utilizing multi-class classification methods for automated sleep disorder prediction. *Information, 15*(8), 426. https://doi.org/10.3390/info15080426.

[7] Farrahi, V., & Rostami, M. (2024). Machine learning in physical activity, sedentary, and sleep behavior research. *Journal of Activity, Sedentary and Sleep Behaviors, 3*, Article 5. https://doi.org/10.1186/s44167-024-00045-9.

[8] Tiwari, S., Arora, D., & Nagar, V. (2022). Supervised approach based sleep disorder detection using non-linear dynamic EEG features. *Measurement: Sensors, 24*, 100469. https://doi.org/10.1016/j.measen.2022.100469.

[9] Chaput, J.-P., Dutil, C., Featherstone, R., et al. (2020). Sleep timing, sleep consistency, and health in adults: A systematic review. *Applied Physiology, Nutrition, and Metabolism, 45*(S1), S232–S247. https://doi.org/10.1139/apnm-2020-0032.

[10] Di Credico, A., Perpetuini, D., Izzicupo, P., et al. (2024). Predicting sleep quality through biofeedback: A machine learning approach using HRV and skin temperature. *Clocks & Sleep, 6*, 322–337. https://doi.org/10.3390/clockssleep6030023.

[11] Crivello, A., Palumbo, F., Barsocchi, P., et al. (2020). Understanding human sleep behaviour by machine learning. *Lecture Notes in Computer Science*. https://doi.org/10.1007/978-3-319-95996-2_11.

[12] Fallmann, S., & Chen, L. (2019). Computational sleep behavior analysis: A survey. *IEEE Access*. https://doi.org/10.1109/ACCESS.2017.

[13] Okano, K., Kaczmarzyk, J. R., Dave, N., et al. (2019). Sleep quality, duration, and consistency are associated with better academic performance. *npj Science of Learning, 4*, Article 16. https://doi.org/10.1038/s41539-019-0055-z.

[14] Phillips, A. J. K., Clerx, W. M., O'Brien, C. S., et al. (2017). Irregular sleep/wake patterns and academic performance. *Scientific Reports, 7*, 3216. https://doi.org/10.1038/s41598-017-03171-4.

[15] Alazaidah, R., Samara, G., Aljaidi, M., et al. (2024). Potential of machine learning for predicting sleep disorders. *Diagnostics, 14*, Article 27. https://doi.org/10.3390/diagnostics14010027.

[16] Greenlund, I. M., & Carter, J. R. (2022). Sympathetic neural responses to sleep disorders. *American Journal of Physiology - Heart and Circulatory Physiology, 322*, H337–H349. https://doi.org/10.1152/ajpheart.00590.2021.

[17] Jones, S. E., Lane, J. M., Wood, A. R., et al. (2019). Genetic studies of accelerometer-based sleep measures. *Nature Communications, 10*, 1585. https://doi.org/10.1038/s41467-019-09576-1.

[18] Yin, J., Jin, X., Shan, Z., et al. (2017). Relationship of sleep duration with mortality and cardiovascular events. *Journal of the American Heart Association, 6*, e005947. https://doi.org/10.1161/JAHA.117.005947.

[19] Sundararajan, K., Georgievska, S., te Lindert, B. H. W., et al. (2021). Sleep classification from wrist-worn accelerometer data. *Scientific Reports, 11*, Article 24. https://doi.org/10.1038/s41598-020-79217-x.

[20] Yuan, H., Plekhanova, T., Walmsley, R., et al. (2024). Self-supervised learning of accelerometer data and sleep-mortality association. *npj Digital Medicine, 7*, Article 86. https://doi.org/10.1038/s41746-024-01065-0.

[21] Perez-Pozuelo, I., Zhai, B., Palotti, J., et al. (2020). The future of sleep health: A data-driven revolution. *npj Digital Medicine, 3*, Article 42. https://doi.org/10.1038/s41746-020-0244-4.

[22] Morokuma, S., Hayashi, T., Kanegae, M., et al. (2023). Deep learning-based sleep stage classification with physiological signals. *Scientific Reports, 13*, 17730. https://doi.org/10.1038/s41598-023-45020-7.

[23] Phillips, A. J. K., Clerx, W. M., O'Brien, C. S., et al. (2017). Irregular sleep/wake patterns and circadian misalignment. *Scientific Reports, 7*, 3216. https://doi.org/10.1038/s41598-017-03171-4.

[24] Sathyanarayana, A., Joty, S., Fernandez-Luque, L., et al. (2016). Sleep quality prediction from wearable data. *JMIR Mhealth Uhealth, 4*(4), e125. https://doi.org/10.2196/mhealth.6562.

[25] Khademi, A., El-Manzalawy, Y., Master, L., et al. (2019). Personalized sleep parameters estimation from actigraphy. *Nature and Science of Sleep, 11*, 387–399. https://doi.org/10.2147/NSS.S220716.

[26] Smith, A., Anand, H., Milosavljevic, S., et al. (2021). Application of machine learning to sleep stage classification. *Proc Int Conf Comput Sci Comput Intell*, 349–354. https://doi.org/10.1109/csci54926.2021.00130.

[27] Mencar, C., Gallo, C., Mantero, M., et al. (2020). Application of machine learning to predict obstructive sleep apnea syndrome. *Health Informatics Journal, 26*(1), 298–317. https://doi.org/10.1177/1460458218824725.

[28] Palotti, J., Mall, R., Aupetit, M., et al. (2019). Benchmark on a large cohort for sleep-wake classification. *npj Digital Medicine, 2*, 50. https://doi.org/10.1038/s41746-019-0126-9.

[29] Wielek, T., Lechinger, J., Wislowska, M., et al. (2018). Sleep in patients with disorders of consciousness using machine learning. *PLOS ONE, 13*(1), e0190458. https://doi.org/10.1371/journal.pone.0190458.

[30] Shiksha Online. (2023, August 9). *XGBoost Algorithm in Machine Learning*. https://www.shiksha.com/online-courses/articles/xgboost-algorithm-in-machine-learning/.

[31] Anshul. (2025, May 1). *Decision Tree Algorithm*. https://www.analyticsvidhya.com/blog/2021/08/decision-tree-algorithm/.

[32] Roshna S H. (2023, April 7). *K-Nearest Neighbors Algorithm*. https://intuitivetutorial.com/2023/04/07/k-nearest-neighbors-algorithm/.

[33] Ajitesh Kumar. (2023, March 27). *Support Vector Machine (SVM) Python Example*. https://vitalflux.com/classification-model-svm-classifier-python-example/.

[34] İlyurek Kılıç. (2023, September 24). *Light GBM: A Powerful Gradient Boosting Algorithm*. https://medium.com/@ilyurek/light-gbm-a-powerful-gradient-boosting-algorithm-fe145a1cd8a6.

[35] Arize, "F1 Score In Machine Learning: What It Is and How To Use It," Arize, Aug. 28, 2024. Available: https://arize.com/blog-course/f1-score/.

[36] GeeksforGeeks. (2025, September 1). Random Forest Algorithm in Machine Learning. Retrieved September 13, 2025, from https://www.geeksforgeeks.org/machine-learning/random-forest-algorithm-in-machine-learning/

[37] Khan, M. M., Masud, M., Aljahdali, S., Kaur, M., & Singh, P. (2021). A comparative analysis of machine learning algorithms to predict Alzheimer's disease. Journal of Healthcare Engineering, 2021(7), Article 9917919. https://doi.org/10.1155/2021/9917919

Addressing Behavioral Patterns of Late-Night Sleepers Using a Supervised Learning Approach

ORIGINALITY REPORT

17%	11%	11%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
2	Submitted to Daffodil International University Student Paper	2%
3	Tayab Uddin Wara, Ababil Hossain Fahad, Adri Shankar Das, Md Mehedi Hasan Shawon. "A systematic review on sleep stage classification and sleep disorder detection using artificial intelligence", Heliyon, 2025 Publication	<1%
4	www.nature.com Internet Source	<1%
5	Michael Elgart, Susan Redline, Tamar Sofer. "Machine and Deep Learning in Molecular and Genetic Aspects of Sleep Research", Neurotherapeutics, 2021 Publication	<1%
6	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	<1%
7	digitalrepository.fccollege.edu.pk Internet Source	<1%
8	Andrea Di Credico, David Perpetuini, Pascal Izzicupo, Giulia Gaggi et al. "Predicting Sleep	<1%