Insomnia Disease Detection Using Machine Learning

By
Rakibul Hasan
ID: 193-15-2935

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised by

Atik Asif Khan Akash

Lecturer
Department of Computer Science and EngineeringDaffodil International University

Co-Supervised by

Mahimul Islam Nadim

Lecturer
Department of Computer Science and EngineeringDaffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

January 12, 2025

APPROVAL

This Project titled **"Insomnia Disease Detection Using Machine Learning"**, submitted by Rakibul Hasan, ID No: **193-15-2935** 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 12/13 January, 2025.

BOARD OF EXAMINERS

Hossain 12.01.2025

Dr. Md. Fokhray Hossain — **Chairman**
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Shah Md Tanvir Siddiquee — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Umaid Hasan — **Internal Examiner**
Lecturer (Senior Scale)
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Nazibur Rahman — **External Examiner**
Technical Lead – Database Administrator
Telenor - Grameen Phone Account

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Atik Asif Khan Akash, Lecturer**, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this projectnor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Akash 12/01/2025

Atik Asif Khan Akash
Lecturer
Department of Computer Science and EngineeringDaffodil International University

Co-Supervised by:

Mahimul Islam Nadim
Lecturer
Department of Computer Science and EngineeringDaffodil International University

Submitted by:

Rakibul Hasan
12-01-2025

Rakibul Hasan
Student ID: 193-15-2935
Department of Computer Science and EngineeringDaffodil 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 as- sisted 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 **Atik Asif Khan Akash** ,**Lecturer**,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** 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

Insomnia happens to be one common sleep disorder that cuts across many people worldwide. It has left untold suffering in his domains that stretch from physical to mental to emotional health problems. Polysomnography and clinical assessment methods are the commonly known traditional ways of diagnosis, but they happen to be time consuming, resource intensive, and often difficult for many to access. This study introduces the application of machine learning (ML) for an automated, efficient, and scalable approach towards insomnia detection. The physiological and behavioral attributes from students of Daffodil International University, which are collected into a dataset through pre- processing and analysis, are used to train various ML models including Gradient Boosting, Random Forest, and Support Vector Machines. The models were tested against evaluation metrics like accuracy, precision, recall, and F1- score. The best model is Gradient Boosting, which achieved testing accuracy of 99.01%, precision of 99%, and F1-score of 99%. Some of the major challenges addressed in this study are imbalanced datasets, complex model interpretability, and ethical considerations such as data privacy. By these results, machine learning turned out to be a feasible option for early detection and accurate latticing of insomnia. Such a system has the potential to revolutionize healthcare by providing accessible, noninvasive, and cost-effective diagnostic tools, thereby improving patient outcomes and advancing the role of ML in sleep medicine.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Insomnia is a common sleep disorder that affects millions of people worldwide, characterized by difficulties in falling asleep, staying asleep, or experiencing restful sleep. Prolonged insomnia can lead to severe health issues, including fatigue, cognitive impairment, emotional distress, and an increased risk of chronic conditions like heart disease, diabetes, and depression. Early detection and diagnosis of insomnia are crucial for timely intervention and treatment [1].

Traditional methods for diagnosing insomnia involve subjective assessments, patient interviews, and sleep studies, which are often time-consuming and resource- intensive. In recent years, the application of machine learning techniques in healthcare has opened new avenues for improving the accuracy and efficiency of disease diagnosis, including sleep disorders like insomnia [2]. Machine learning algorithms can analyze large volumes of data, such as sleep patterns, physiological signals, and behavioral indicators, to identify key features associated with insomnia and predict its occurrence with greater precision.

Insomnia among Bangladeshi students, particularly at the university level, is a growing concern that has been exacerbated by various academic, societal, and technological factors. Research reveals that approximately 76.5% of university students experience moderate to severe insomnia, closely linked to mental health issues such as anxiety and depression. Factors like prolonged screen time, disrupted sleep schedules, and heightened academic pressures significantly contribute to this prevalence. A survey of 1,640 students from various universities and colleges in Bangladesh highlighted that nearly 74.5% of students face difficulty due to unmet psychological needs. Moreover, 73.8% report a decline in self-confidence due to insomnia and related academic challenges [3].

This study explores the use of machine learning models for insomnia detection, leveraging diverse data sources, such as sleep duration, heart rate variability, and daily activity levels, to develop an automated and scalable solution for diagnosing insomnia. By employing advanced algorithms and data-driven approaches, this research aims to improve the early detection of insomnia, leading to more effective treatment and better overall patient outcomes.

1.2 Motivation

This sleep disorder, known as insomnia, is common around the globe and has grave effects on people; it upsets all three areas of health: mental, emotional, and physical. If left untreated or unnoticed, the effects can be considerable, including chronic fatigue, depletion of immune response, cardiovascular diseases, as well as mental health-related diseases like anxiety and depression. Thus, a timely diagnosis of insomnia will help avert these effects, thereby ensuring improved quality of life among those impacted.

With machine learners, it has become possible to automate, speed up, and improve diagnostic standards. Such algorithms are trained with millions of patient data and can thus derive correlations and patterns undetectable through previous means. Such systems can reveal early and accurate detection from sleep pattern analysis, physiological parameters, and other lifestyle factors.

The potential of machine learning for insomnia detection lies in its ability to:

Improve Early Intervention: Using machine learning applied to data analysis to pick up any signs of insomnia earlier, allowing for better and earlier intervention.

Be Accessible: Have detection methods which are automated systems that potentially allow the diagnosis tool to be much more widely available, even in remote or underserved areas.

Improve Accuracy: More available ML could mean less human error and subjective bias in diagnosis, generating more reliable results.

Offer Customized Interventions: Advanced ML models could be able to consider individual differences in sleep patterns and provide better recommendations for intervention on an individual basis.

Reduce Cost and Time: Automated systems can reduce decreased time and costs associated with traditional diagnostics significantly.

This application of machine learning is aimed to break the current methods used to diagnose and manage insomnia; thereby changing the lives of numerous individuals traumatized by the ailment.

1.3 Objectives

The primary objectives of this study are as follows:

1. Develop a Machine Learning Model for Insomnia Detection: To design and implement a machine learning-based model capable of analyzing physiological data such as sleep patterns, heart rate, and activity levels to accurately detect insomnia.

2. Identify Key Features Associated with Insomnia: To determine the most relevant physiological and behavioral features that contribute to insomnia detection, using data

collected from wearable devices or clinical settings.

3. Improve Diagnostic Accuracy and Efficiency: To enhance the accuracy and efficiency of insomnia diagnosis by comparing various machine learning algorithms (e.g., decision trees, random forests, neural networks) and identifying the most effective model for detecting insomnia.

4. Evaluate Model Generalization and Robustness: To ensure the developed model can generalize well across diverse populations and handle noisy, incomplete, or heterogeneous data, leading to a reliable solution for insomnia detection.

5. Provide a Scalable and Non-Invasive Diagnostic Tool: To create a scalable, non-invasive, and cost-effective diagnostic tool for healthcare providers, enabling continuous monitoring and early detection of insomnia outside traditional clinical environments.

6. Contribute to Personalized Treatment: To facilitate more personalized treatment plans for insomnia patients by providing healthcare professionals with detailed insights derived from machine learning analysis of patient data.

1.4 Methodology

The methodology for developing an insomnia detection system using machine learning involves a systematic approach to data collection, preprocessing, model development, and evaluation. Below are the key steps involved:

1. Problem Definition and Objective Setting
Clearly define the scope of the project: detect insomnia based on physiological, behavioral, and lifestyle data.

2. Data Collection

- Data Sources:
    - Publicly available sleep datasets from sources like the National Sleep Research Resource (NSRR) or Kaggle.
    - Physiological data from wearable devices (e.g., heart rate, movement, oxygen levels).
    - Self-reported data from surveys or sleep diaries (e.g., sleep duration, quality, and patterns).
- Sample Size: Collect a diverse and sufficiently large dataset to ensure the model generalizes well.
- Ethical Considerations: Ensure that data collection complies with privacy and ethical standards, obtaining necessary consents.

3. Data Preprocessing

- Data Cleaning:
  Remove incomplete or inconsistent records.
  Handle missing values using imputation techniques.
- Normalization and Scaling:
  Normalize continuous variables like heart rate or sleep duration to a common scale.
- Data Labeling:
  Categorize data into labels such as "Insomniac" and "Non-Insomniac" based on diagnostic criteria (e.g., ICD-10 or DSM-5).
- Train-Test Split:
  Divide the dataset into training, validation, and test sets (e.g., 70-20-10 split).

4. Machine Learning Model Selection
   - Supervised Learning: For labeled data, consider models like Random Forests, Support Vector Machines (SVM), Gradient Boosting (e.g., XGBoost), or Neural Networks.
   - Deep Learning: For time-series or complex datasets, use architectures like Long Short-Term Memory (LSTM) networks or Convolutional Neural Networks (CNNs).
   - Consider ensemble methods to improve performance.

5. Model Training
   - Train the model using the training dataset.
   - Apply techniques like cross-validation to avoid overfitting.

6. Model Evaluation
   - Accuracy: Overall correctness of predictions.
   - Precision and Recall: Effectiveness in identifying insomniacs.
   - F1-Score: Balance between precision and recall.
   - ROC-AUC Score: Ability to distinguish between insomniac and non-insomniac cases.

Analyze confusion matrix to identify misclassifications and refine the model if necessary.

7. Documentation and Reporting
   - Document the entire methodology, including data sources, preprocessing steps, model architecture, and evaluation results.
   - Publish findings in journals or present them at conferences to contribute to the field.

By following this methodology, the project ensures a comprehensive and scientific approach to building a reliable, accurate, and practical insomnia detection system using machine learning.

1.5 Project Outcome

The project aims to develop an efficient, accurate, and user-friendly machine learning-based system for detecting insomnia. The outcomes of this project are expected to bring significant advancements in the diagnosis and management of insomnia, benefiting both patients and healthcare providers. The anticipated outcomes include:

1. Development of a Machine Learning Model: A robust machine learning model capable of analyzing sleep-related data (e.g., sleep duration, quality, disruptions, physiological parameters) to detect patterns indicative of insomnia.
2. Automated Insomnia Detection System: A system that can process input data from wearable devices, mobile health apps, or patient self-reports to automatically diagnose insomnia.
3. Improved Diagnosis Accuracy: Enhanced precision and reliability compared to traditional diagnostic methods through the use of data-driven insights and advanced algorithms.
4. Healthcare Integration: A tool that healthcare providers can integrate into their workflows to support clinical decision-making, enabling faster and more informed diagnoses.

By achieving these outcomes, the project will not only enhance the early detection and management of insomnia but also pave the way for future advancements in applying machine learning to other sleep-related and mental health disorders.

1.6 Organization of the Report

This report is structured to provide a comprehensive understanding of the project "Insomnia Disease Detection using Machine Learning," detailing every aspect from motivation to implementation. The organization is as follows:

Chapter 1. Introduction: This chapter introduces the project, including the background, problem statement, motivation, objectives, and a brief overview of the methodology. It also outlines the anticipated outcomes and the structure of the report.

Chapter 2. Background: This section provides the foundational knowledge necessary to understand the project, including a detailed literature review, related research studies, and a gap analysis to highlight the need for this work.

Chapter 3. Research Methodology: This chapter describes the methodology adopted for the project, detailing the data collection process, feature engineering, model selection, and evaluation criteria. It also includes system design, task allocation, and the proposed project plan.

Chapter 4. Implementation and Results: This chapter focuses on the implementation details, including the environment setup, performance evaluation, and comparative analysis of the machine learning models used. It also discusses the results obtained and their implications.

Chapter 5. Engineering Standards and Design Challenges: This section discusses the compliance with relevant software and hardware standards, the challenges encountered during the project, and the impact on society, environment, and sustainability. Ethical considerations and the sustainability plan are also elaborated. Chapter 6. Conclusion: The concluding chapter summarizes the project, highlights its limitations, and proposes potential areas for future work.

Chapter 2

Background

2.1 Introduction

The increasing prevalence of insomnia as a global health concern has spurred significant research efforts to better understand, diagnose, and manage this debilitating sleep disorder. Traditional methods for diagnosing insomnia, such as polysomnography (PSG) and clinical assessments, are resource-intensive, time-consuming, and often inaccessible to the broader population. This has created a pressing need for innovative solutions that are both efficient and scalable.

Machine learning (ML), with its ability to analyze complex patterns in large datasets, has emerged as a promising approach in healthcare, particularly for early diagnosis and decision-making processes. Over the past decade, researchers have explored the application of ML techniques in sleep medicine to automate the detection of insomnia and other sleep-related disorders. These studies aim to harness the predictive capabilities of ML models to improve diagnostic accuracy, reduce costs, and enhance accessibility to care.

This literature review examines the current state of research in insomnia detection using machine learning. It provides an overview of the methodologies, datasets, and ML techniques employed in previous studies, highlights their achievements and limitations, and identifies gaps in the existing knowledge. By understanding the progress and challenges in this field, this review sets the foundation for the development of an advanced and effective ML-based system for insomnia detection.

2.2 Literature Review

Calderon et al. [1] this study investigated the relationships between insomnia and various mental disorders among U.S. college students. Utilizing an elastic net regularization model, the researchers assessed the associations between insomnia severity and seven mental disorders, including major depressive disorder (MDD), generalized anxiety disorder, and post-traumatic stress disorder. The findings revealed that MDD had the most significant association with increased insomnia severity. Further analysis using Bayesian network methods suggested that symptoms such as depressed mood, fatigue, and low self-esteem may precede the development of insomnia. These results underscore the intricate link between depression and insomnia in college students, highlighting the importance of addressing depressive symptoms to mitigate sleep disturbances in this population.

M. Ingle et al. [2] in their paper comprehensively examine the advancements in utilizing machine learning (ML) and deep learning (DL) algorithms for the automated detection of insomnia through physiological data analysis. The authors assess various ML and DL models, highlighting their respective strengths and limitations in identifying insomnia-related patterns. They also discuss the challenges associated with data quality, feature extraction, and model interpretability. The review underscores the necessity for further research to enhance the accuracy and reliability of these automated systems, aiming to integrate them effectively into clinical practice for improved insomnia diagnosis and management.

M. M. Islam *et al* [4] this paper proposes a machine learning model to predict chronic insomnia using external symptoms. The authors compare seven classifiers, with logistic regression achieving the best accuracy (98%). The study highlights the importance of addressing insomnia early due to its links to other health issues. The model was trained on a custom dataset and demonstrated strong performance through cross-validation, offering a promising tool for insomnia detection.

C. Jansen *et al* [5] this paper introduces a two-stage method to automatically detect insomnia from EEG recordings. In the first stage, sleep stages and insomnia are classified at the epoch level using DNN-based models. The second stage extracts subject- level features and applies machine learning classifiers like SVM, LDA, and CART to differentiate between control and insomnia subjects. The model achieved an F1 score of 0.88 with an SVM classifier.

Y. Furukawa *et al* [6] the study examines the relationship between cognitive, emotional, and metacognitive variables in predicting insomnia severity among college students. The findings highlight the importance of cognitive arousal, dysfunctional beliefs about sleep, and depressive symptoms as significant predictors of insomnia. The results suggest that cognitive factors may play a more significant role than emotional variables.

C. Baglioni *et al* [7] this paper investigates sleep quality using wearable actigraphy sensors that analyze physical activity and sleep patterns. Machine learning models such as CNN, MLP, and LSTM-RNN were employed to predict sleep efficiency, with CNN outperforming other methods. The study highlights the potential of wearable devices in non-invasive insomnia detection.

R. Alazaidah *et al* [8] this study focuses on detecting psychological stress among university students, which is closely associated with insomnia. It utilizes classifiers such as SVM, NB, LR, and RF, with SVM achieving the best performance at 85.71% accuracy. The research underscores the connection between stress and sleep disorders like insomnia.

H. Doos Ali Vand *et al* [9] this research predicts sleep disorders in an asthma cohort using machine learning techniques like SVM, KNN, and RF, along with deep learning models such as CNN and RNN. CNN achieved the highest accuracy of 95.1%, outperforming traditional methods. The study demonstrates the effectiveness of deep learning in detecting sleep disorders.

M. Shahin *et al* [10] this study compares the effectiveness of CNN, LSTM, and GRU in predicting sleep disorders from large datasets. CNN emerged as the best- performing model, with accuracy rates higher than traditional machine learning methods. The paper emphasizes the scalability of deep learning techniques for large- scale sleep disorder detection.

S. Li *et al* [11] this paper compares different machine learning algorithms such as SVM, RF, and KNN for insomnia prediction, with SVM achieving the highest accuracy (92%). The study uses a dataset of 100 patients and identifies mobility and vision problems as significant factors influencing insomnia.

Q. Lu *et al* [12] this paper discusses the relationship between insomnia and suicide ideation in U.S. Army personnel. It highlights the role of insomnia as a predictor for suicide-related behaviors, suggesting the need for targeted interventions to address sleep disorders in high-risk populations.

T. Roth [13] This study focuses on classifying depression levels using machine learning, which is closely associated with insomnia. It uses speech, text, and image data to predict depression severity, applying classifiers like CNN and SVM. The paper demonstrates the interconnectedness of mental health disorders like depression and insomnia.

Table 2.1: Summary of Literature Reviewed.

Reference	**Dataset/Method**	**Models Used**	**Best Model and Performance**	**Key Findings**
[1] Islam et al. (2020)	Custom dataset for chronic insomnia prediction	Logistic Regression, others	98% (Logistic Regression)	Logistic regression achieved the best accuracy; emphasizes early detection of insomnia due to its link with other health issues.
[2] Jansen et al. (2019)	EEG recordings for insomnia detection	DNN-based models, SVM, LDA, CART	0.88 (DNN)	Two-stage method; DNN for sleep stage classification, followed by machine learning models for insomnia detection.

[3] Baglioni et al. (2011)	Actigraphy and physical activity data	CNN, MLP, LSTM-RNN	CNN outperforms other models	Wearable actigraphy sensors analyzed physical activity and sleep patterns; CNN showed the best performance for sleep efficiency.
[4] Alazaidah et al. (2023)	Psychological stress among university students	SVM, Naïve Bayes, LR, RF	85.71% (SVM)	SVM achieved best performance in detecting stress, which is closely linked to insomnia.
[5] Vand et al. (2023)	Asthma cohort for sleep disorder prediction	SVM, KNN, RF, CNN, RNN	95.1% (CNN)	Deep learning models, particularly CNN, outperformed traditional methods in predicting sleep disorders in asthma patients.
[6] Shahin et al. (2018)	Large dataset for insomnia classification	CNN, LSTM, GRU	CNN highest	CNN outperformed LSTM and GRU, showing deep learning's effectiveness for large-scale sleep disorder detection.
[7] Li et al. (2023)	100-patient dataset for insomnia severity	SVM, RF, KNN	92% (SVM)	SVM achieved the highest accuracy in predicting insomnia severity, with mobility and vision problems as influencing factors.

2.3 Gap Analysis

Recent advancements in machine learning (ML) have significantly enhanced the automatic identification of insomnia through the analysis of physiological signals. A systematic review by Ingle et al. (2024) [3] highlights the progress in this area, noting the development of various ML and deep learning (DL) algorithms trained on annotated physiological data. However, the study also identifies a notable research gap, emphasizing the need for improved accuracy and reliability in current detection methods.

Additionally, Calderon et al. (2024) [2] utilized supervised ML and Bayesian network analyses to explore associations between insomnia and mental disorders among college students. Their findings suggest a complex interplay between insomnia and conditions like major depressive disorder, underscoring the necessity for further research to understand these relationships fully.

These studies collectively indicate that while ML techniques have advanced insomnia detection, challenges persist in achieving high accuracy and understanding the disorder's multifaceted nature. Addressing these gaps is crucial for developing more effective and reliable ML-based diagnostic tools for insomnia.

2.4 Summary

The literature review highlights the growing interest in leveraging machine learning for insomnia detection as a response to the limitations of traditional diagnostic methods. Studies reviewed in this segment demonstrate significant progress in using machine learning to analyze sleep-related data, including physiological signals, behavioral patterns, and self-reported sleep metrics. Various machine learning models, such as Support Vector Machines (SVM), Random Forests, Gradient Boosting, and Neural Networks, have been employed with notable success in improving diagnostic accuracy and efficiency.

Several studies utilize data from wearable devices, mobile health applications, and publicly available sleep datasets, showcasing the potential for real-world implementation and scalability. Feature engineering and data preprocessing have been critical in extracting meaningful insights, while advanced algorithms have proven effective in distinguishing insomniac individuals from non-insomniacs. Despite these achievements, challenges remain, such as handling imbalanced datasets, ensuring interpretability of complex models, and addressing ethical concerns regarding data privacy and bias.

The review identifies gaps in existing research, including the need for standardized datasets, real-time detection capabilities, and broader generalizability across diverse populations. These insights provide a clear direction for future work, emphasizing the development of robust, accessible, and clinically validated machine learning systems for insomnia detection.

Chapter 3

Research Methodology

3.1 Methodology

3.1.1 Overview

The methodology for developing a machine learning-based system for insomnia detection involves a structured and systematic approach to ensure accuracy, efficiency, and reliability. It begins with defining the problem and objectives, followed by collecting and preprocessing data from diverse sources such as wearable devices, sleep diaries, and publicly available datasets. Feature engineering and selection are employed to identify key indicators of insomnia, which are then used to train machine learning models. These models are evaluated using metrics like accuracy, precision, recall, and F1-score to ensure their effectiveness in detecting insomnia. The methodology also includes steps for integrating the system into user-friendly platforms, testing its real-world performance, and ensuring ethical considerations such as data privacy and fairness. By following this approach, the project aims to create a robust, scalable, and impactful solution for early and accurate insomnia detection.

3.1.2 Proposed Methodology

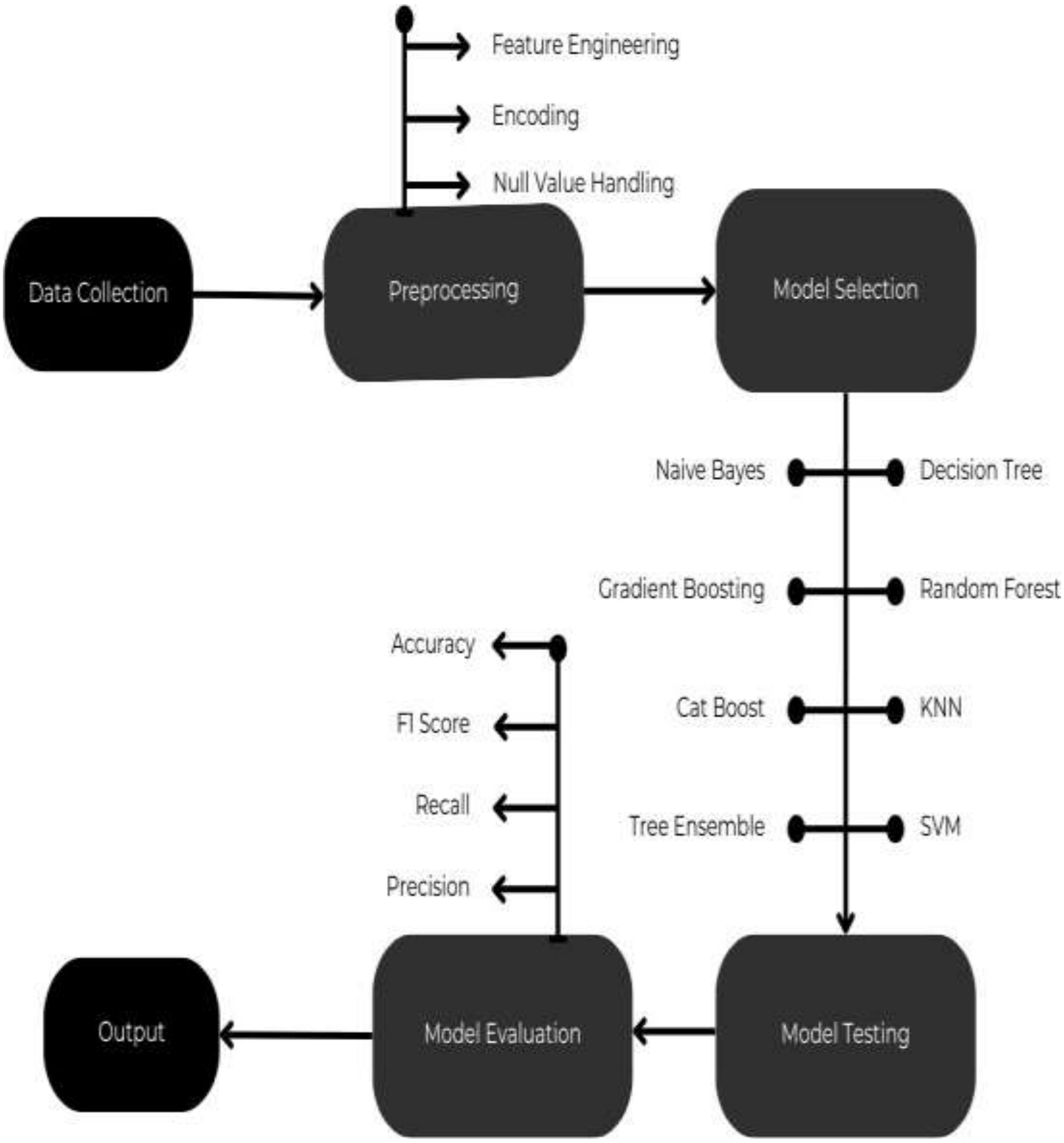


Figure 3.1: Proposed Methodology

3.2 Detailed Methodology and Design

3.2.1 Data Collection: Data collection is the most crucial part for this research. We used fully raw dataset collected by us. We collected data from students of Daffodil International University from March to July. In this dataset, there were 20 attributes and 1020 instances. In the target column, there were 2 classes Yes & No, where —Yes|| were 798 and —No|| were 222.

Age	Gender	fall_asleep_	Awakening	awakening_	Total_sleep	quality_sle	well_being	Functioning	Sleepiness_	marital_sta	pressure/st	depressed	smoke	alcohol	tea_coffee	addicted_g	phone_bef	study work	illness
22	Male	0: No proble	0: No proble	1: A little ea	1: Slightly in	1: Slightly u	0: Normal	0: Normal	0: None	Single	Yes	No	No	No	No	No	Yes	Yes	No
25	Male	0: No proble	1: Minor prol	0: Not earlie	0: Sufficient	0: Satisfacto	0: Normal	0: Normal	2: Consideral	Single	No	No	No	No	No	Yes	Yes	No	No
22	Male	0: No proble	2: Consideral	1: A little ea	1: Slightly in	0: Satisfacto	0: Normal	0: Normal	1: Mild	Single	Yes	No	No	No	Yes	No	Yes	Yes	No
20	Male	0: No proble	0: No proble	0: Not earlie	0: Sufficient	0: Satisfacto	0: Normal	0: Normal	0: None	Single	No	No	No	No	Yes	Yes	Yes	Yes	No
24	Male	0: No proble	0: No proble	1: A little ea	1: Slightly in	1: Slightly u	0: Normal	0: Normal	0: None	Single	Yes	Yes	No	No	Yes	No	Yes	Yes	No
21	Male	0: No proble	0: No proble	0: Not earlie	3: Very insuf	3: Very unsa	3: Very decr	3: Very decr	1: Mild	Single	No	No	No	No	Yes	No	Yes	Yes	No
19	Male	0: No proble	0: No proble	1: A little ea	0: Sufficient	0: Satisfacto	0: Normal	0: Normal	2: Consideral	Single	No	No	No	No	Yes	Yes	Yes	No	No
21	Male	0: No proble	0: No proble	1: A little ea	0: Sufficient	0: Satisfacto	0: Normal	0: Normal	2: Consideral	Single	No	No	No	No	Yes	No	No	No	No
21	Male	0: No proble	2: Consideral	0: Not earlie	1: Slightly in	1: Slightly u	1: Slightly de	1: Slightly de	1: Mild	Single	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No
23	Male	0: No proble	1: Minor prol	0: Not earlie	2: Markedly i	1: Slightly u	1: Slightly de	1: Slightly de	2: Consideral	Single	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No
21	Male	0: No proble	2: Consideral	1: A little ea	1: Slightly in	1: Slightly u	0: Normal	0: Normal	2: Consideral	Single	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes
21	Male	0: No proble	1: Minor prol	1: A little ea	1: Slightly in	1: Slightly u	2: Markedly	2: Markedly	0: None	Single	No	No	No	No	Yes	No	Yes	No	No
22	Male	0: No proble	1: Minor prol	2: Markedly	0: Sufficient	0: Satisfacto	1: Slightly de	1: Slightly de	1: Mild	Single	No	No	No	No	Yes	No	Yes	No	No
20	Male	0: No proble	1: Minor prol	0: Not earlie	0: Sufficient	0: Satisfacto	0: Normal	0: Normal	0: None	Single	Yes	Yes	No	No	No	Yes	Yes	Yes	No
24	Male	0: No proble	1: Minor prol	0: Not earlie	0: Sufficient	0: Satisfacto	0: Normal	0: Normal	0: None	Single	No	No	Yes	No	Yes	Yes	Yes	No	No
20	Male	0: No proble	1: Minor prol	0: Not earlie	1: Slightly in	0: Satisfacto	0: Normal	0: Normal	0: None	Single	No	No	No	No	Yes	No	No	No	No
20	Male	0: No proble	2: Consideral	1: A little ea	1: Slightly in	2: Markedly	1: Slightly de	1: Slightly de	1: Mild	Single	No	Yes	No	No	Yes	No	Yes	No	No
23	Male	0: No proble	1: Minor prol	1: A little ea	1: Slightly in	0: Satisfacto	1: Slightly de	1: Slightly de	1: Mild	Single	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
23	Male	0: No proble	0: No proble	2: Markedly	1: Slightly in	1: Slightly u	1: Slightly de	1: Slightly de	0: None	Single	No	No	No	No	Yes	No	Yes	No	No
20	Male	0: No proble	0: No proble	0: Not earlie	0: Sufficient	0: Satisfacto	0: Normal	0: Normal	3: Intense	Single	Yes	Yes	No	No	Yes	Yes	Yes	No	No
22	Male	0: No proble	1: Minor prol	2: Markedly	1: Slightly in	1: Slightly u	2: Markedly	2: Markedly	2: Consideral	Single	Yes	No	No	No	Yes	No	Yes	Yes	No
23	Male	0: No proble	0: No proble	1: A little ea	0: Sufficient	0: Satisfacto	0: Normal	0: Normal	1: Mild	Single	No	No	No	No	Yes	No	Yes	No	Yes
20	Male	0: No proble	0: No proble	0: Not earlie	1: Slightly in	1: Slightly u	0: Normal	0: Normal	1: Mild	Single	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes
20	Male	0: No proble	0: No proble	1: A little ea	1: Slightly in	1: Slightly u	1: Slightly de	1: Slightly de	1: Mild	Single	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
19	Male	0: No proble	0: No proble	1: A little ea	0: Sufficient	0: Satisfacto	0: Normal	0: Normal	0: None	Single	No	No	Yes	No	Yes	No	Yes	No	No
19	Male	0: No proble	0: No proble	0: Not earlie	0: Sufficient	0: Satisfacto	0: Normal	0: Normal	0: None	Single	No	No	No	No	Yes	Yes	Yes	No	Yes
22	Male	0: No proble	0: No proble	0: Not earlie	2: Markedly i	1: Slightly u	1: Slightly de	1: Slightly de	1: Mild	Single	Yes	No	No	No	Yes	No	Yes	Yes	No

Figure 2. Sample Dataset

3.2.2 Data Preprocessing & Feature Engineering: At first, we checked for the null values. There were no null values. Most of the attributes were string. So, we had to convert them in numeric value. For some attributes we used replace techniques and for rest of the attributes we used encoding techniques. For encoding we used —Label Encoder||.

For feature engineering we used —SelectKBest|| techniques. On the basis of feature engineering result, we used best 10 columns for this research.

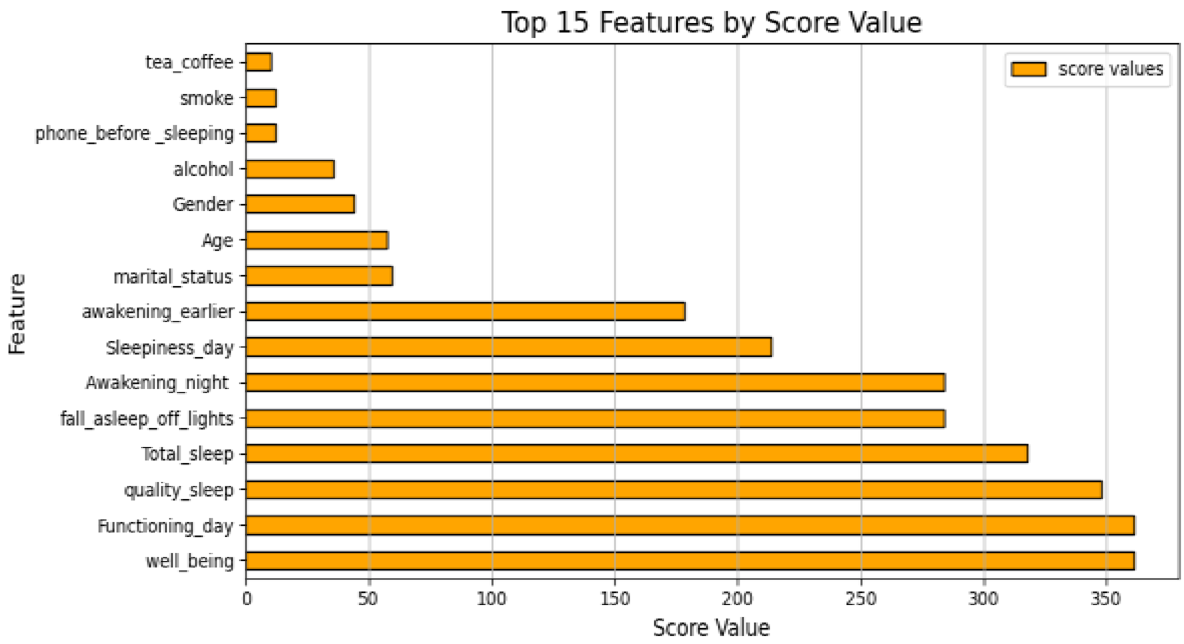


Figure 3. Result of Feature Engineering

3.2.3 Model Selection:

1. Decision Tree

- **Description**: A tree-structured algorithm where each internal node represents a feature or attribute, each branch represents a decision rule, and each leaf represents an outcome.
- **Strengths**: Easy to interpret, handle both numerical and categorical data, and requires minimal data preprocessing.
- **Limitations**: Prone to overfitting, especially with complex datasets, unless pruned or regularized.

2. Random Forest

- **Description**: An ensemble method that builds multiple decision trees during training and combines their outputs (via averaging for regression or majority voting for classification).
- **Strengths**: Handles overfitting better than individual decision trees, is robust to noise, and works well with large datasets.
- **Limitations**: Computationally expensive due to the generation of multiple trees and may be less interpretable than a single decision tree.

3. K-Nearest Neighbors (KNN)

- **Description**: A lazy learning algorithm that assigns a class to a data point based on the majority vote of its k-nearest neighbors in the feature space.
- **Strengths**: Simple to implement, works well for smaller datasets, and makes no assumptions about data distribution.

- **Limitations**: Computationally intensive for large datasets and sensitive to the choice of k and feature scaling.

4. Naive Bayes

- **Description**: A probabilistic classifier based on Bayes' theorem, assuming independence among features.
- **Strengths**: Fast to train, effective for high-dimensional datasets, and performs well with categorical data and text classification.
- **Limitations**: The independence assumption is often unrealistic, and it may perform poorly on datasets with correlated features.

5. Support Vector Machine (SVM)

- **Description**: A supervised learning algorithm that finds the optimal hyperplane to separate data points in feature space.
- **Strengths**: Effective for high-dimensional spaces and works well with both linear and nonlinear boundaries (using kernels).
- **Limitations**: Computationally intensive for large datasets and sensitive to the choice of kernel parameters.

6. XGBoost

- **Description**: A gradient-boosting framework that uses decision trees as weak learners and optimizes model performance iteratively.
- **Strengths**: Highly efficient, handles missing data, works well with structured/tabular data, and offers features for regularization to avoid overfitting.
- **Limitations**: Can be computationally intensive and requires careful tuning of hyperparameters.

7. Gradient Boosting

- **Description**: An ensemble method that builds models sequentially, with each model correcting errors of the previous one.
- **Strengths**: Robust to overfitting with proper tuning, capable of handling non-linear relationships, and performs well on structured data.
- **Limitations**: Training can be slow, and the method is sensitive to hyperparameter choices.

8. CatBoost

- **Description**: A gradient boosting framework optimized for categorical data, automatically handling categorical features without requiring one-hot encoding.
- **Strengths**: Efficient handling of categorical data, avoids overfitting, and supports GPU acceleration for faster training.
- **Limitations**: Requires significant computational resources for large datasets.

9. Tree Ensemble

- **Description**: An ensemble technique that combines multiple decision trees to improve accuracy and robustness, often implemented with gradient boosting or random forests.
- **Strengths**: High accuracy, robust to overfitting, and capable of handling complex relationships in the data.
- **Limitations**: Computationally intensive and less interpretable compared to individual decision trees.

These algorithms offer diverse strengths, making them suitable for different types of datasets and problems. Ensemble methods like Random Forest, XGBoost, CatBoost, and Tree Ensemble often outperform individual algorithms, especially on complex datasets, due to their ability to reduce bias and variance.

3.2.4 Model Evaluation:

The models will be evaluated using a comprehensive set of metrics, including precision, recall, F1-score, and accuracy. These metrics provide a detailed assessment of each model's performance in classifying diseases:

Precision: Measures the proportion of correctly predicted disease cases out of all predicted cases, indicating the model's accuracy in making positive predictions.

$$Precision = \frac{True\ Positives}{TruePositives + FalsePositives} \quad (1)$$

Recall: Also known as sensitivity, this metric measures the proportion of actual disease cases that were correctly identified by the model.

$$Recall = \frac{True\ Positives}{TruePositives + False\ Negetives} \quad (2)$$

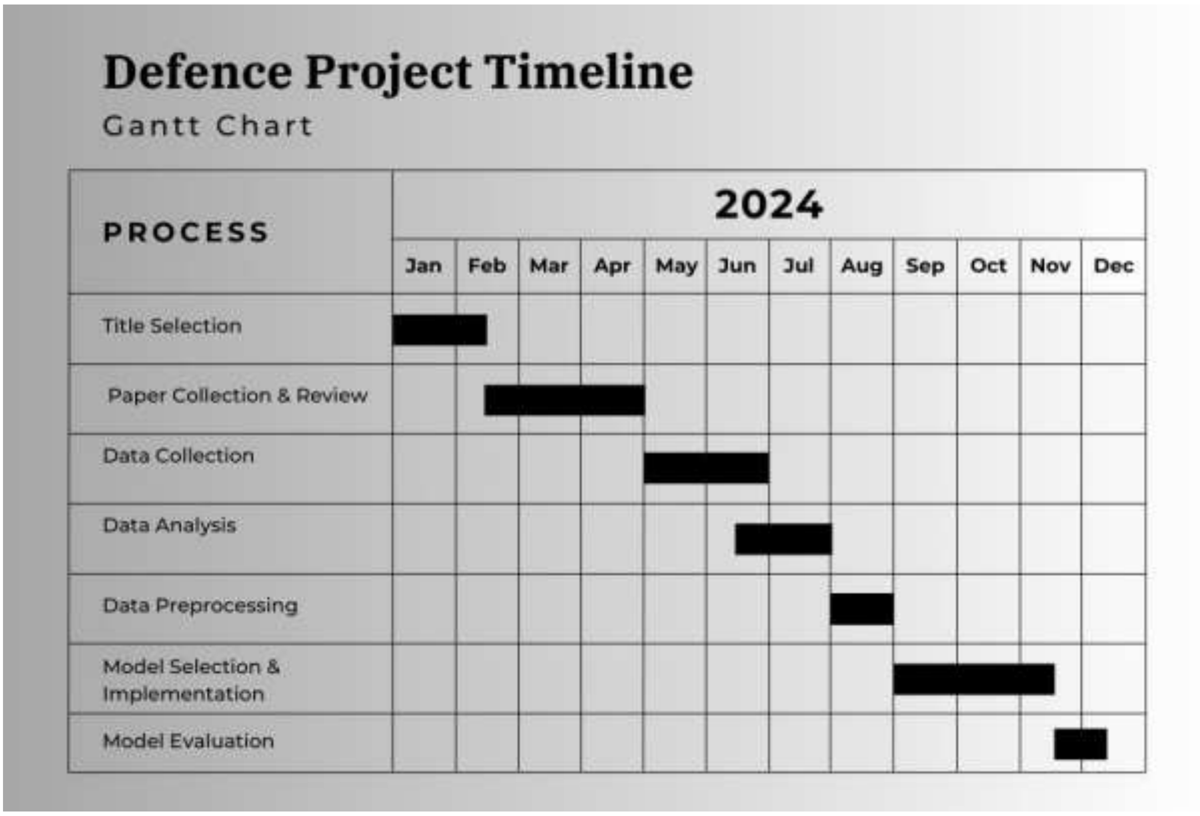
F1-Score: Combines precision and recall into a single metric by calculating their harmonic mean, offering a balanced view of the model's performance.

$$F1 = 2 * \frac{Precesion\ and\ Recall}{TruePositives + FalsePositives} \quad (3)$$

Accuracy: Represents the overall correctness of the model by calculating the proportion of correct predictions out of all predictions made.

3.3 Project Plan

Planning for any project is very essential. The duration for this project was 1 year. So, planning was very important for successfully complete this project. The project timeline is given below:



3.4 Task Allocation

There is no member in my team. So, all the tasks are completed of my own.

3.5 Summary

The methodology for insomnia disease detection using machine learning follows a structured approach, encompassing problem definition, data collection, model development, and evaluation. Data is gathered from various sources, including wearable devices, sleep diaries, and publicly available datasets, and is carefully preprocessed through cleaning, normalization, and feature engineering to ensure relevance and quality. Machine learning models, such as Random Forests, SVM, and Neural Networks, are trained and validated using diverse datasets, with performance evaluated through metrics like accuracy, precision, recall, and F1-score. Ethical considerations, including data privacy and bias mitigation, are integral throughout the process. The system is designed for real-world application, with a focus on accessibility and scalability, ensuring it can be integrated into user-friendly platforms for widespread adoption. This comprehensive methodology ensures the development of an accurate, efficient, and ethical solution for insomnia detection.

Chapter 4

Implementation and Results

4.1 Environment Setup

4.2 Comparative Analysis

For this research, we used total 9 algorithms among them some were traditional algorithms and some were boosting algorithms and one was Ensemble Method.

Table 4.1. Training and Testing Score of Each Algorithm

Algorithm	Training Score	Testing Score
Decision Tree	100	92.65
Random Forest	100	95.09
KNN	98.04	95.09
Naïve Bayes	97.03	92.15
SVM	96.56	92.15
XGBoost	100	95.588
Gradient Boosting	100	99.01
Cat Boost	98.52	95.09
Tree Ensemble	100	95.09

The Table 4.1 presents the performance of various machine learning algorithms for insomnia disease detection, highlighting both training and testing scores. Several algorithms, such as Decision Tree, Random Forest, XGBoost, Gradient Boosting, and Tree Ensemble, show perfect or near-perfect training scores (100%), indicating potential overfitting. In terms of testing performance, Gradient Boosting stands out with the highest testing score of 99.01%, suggesting it generalizes the best to new data. Random Forest, XGBoost, and Tree Ensemble follow closely with testing scores of 95.09%, while KNN and CatBoost achieve 95.09% testing scores but with slightly lower training scores, indicating good generalization ability but some overfitting. Naïve Bayes and SVM perform the weakest with testing scores of 92.15%, showing lower generalization capabilities. Overall, Gradient Boosting offers the best balance of training and testing performance, making it the most suitable algorithm for this task, though Random Forest and XGBoost are also strong choices. If avoiding overfitting is a priority, KNN and CatBoost provide solid alternatives, albeit with marginally lower performance.

Table 4.2. Performance of Each Algorithm

Algorithm	**Precision**	**Recall**	**F1 Score**
Decision Tree	0.91	0.88	0.90
Random Forest	0.95	0.91	0.93
KNN	0.94	0.92	0.93
Naïve Bayes	0.92	0.86	0.88
SVM	0.95	0.84	0.88
XGBoost	0.96	0.92	0.94
Gradient Boosting	0.99	0.98	0.99
Cat Boost	0.96	0.90	0.93
Tree Ensemble	0.97	0.90	0.93

The performance metrics for the insomnia disease detection models show that Gradient Boosting outperforms all other algorithms, achieving the highest accuracy (99%), precision (99%), recall (98%), and F1 score (99%), making it the most reliable model for this task. XGBoost also performs excellently with 96% accuracy, 94% F1 score, and strong precision and recall, while Random Forest, KNN, and CatBoost follow closely, offering 95% accuracy and 93% F1 scores, demonstrating their effectiveness as well. Decision Tree provides good results with 93% accuracy, but its performance is slightly lower in recall and F1 score compared to the others. Naïve Bayes and SVM are less effective, particularly in terms of recall and F1 score, with SVM showing the weakest recall at 84%. Overall, Gradient Boosting is the best choice for this task due to its superior balance of precision, recall, and F1 score, while XGBoost and Random Forest are also strong contenders.

We can see the clear performances of the models from the confusion matrix shown below in Figure 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, & 4.9. Also, the ROC curve mention below in Figure 4.10, 4.11, 4.12, 4.13, 4.14, 4.15, 4.16, 4.17, & 4.18 can tell the story.

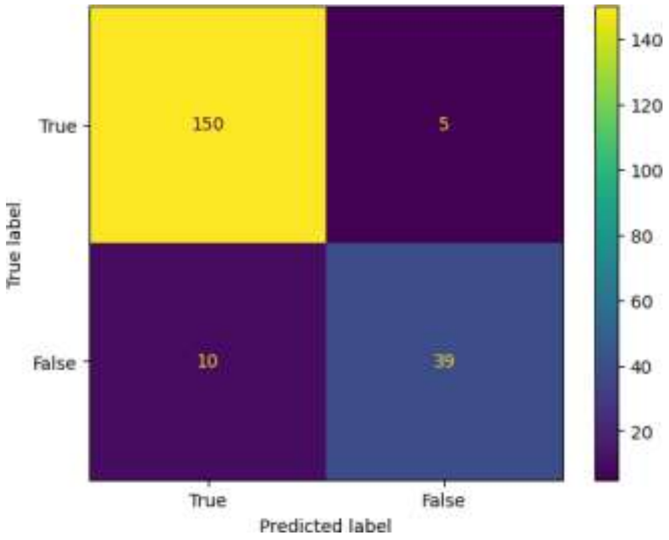


Figure 4.1. Confusion Matrix of Decision Tree

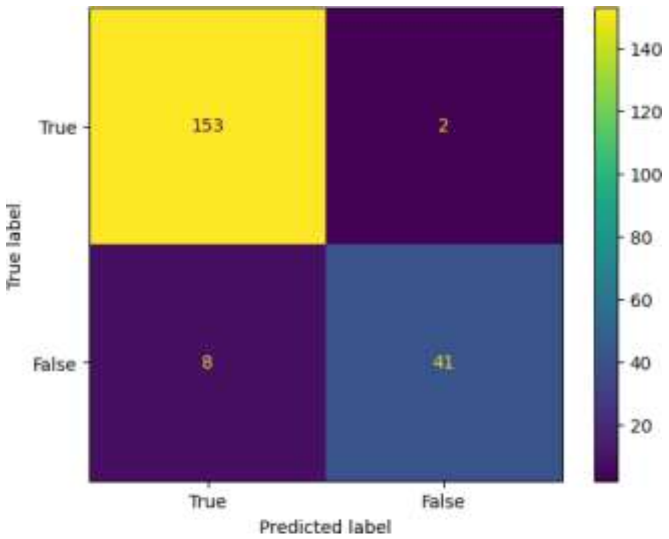


Figure 4.2. Confusion Matrix of Random Forest

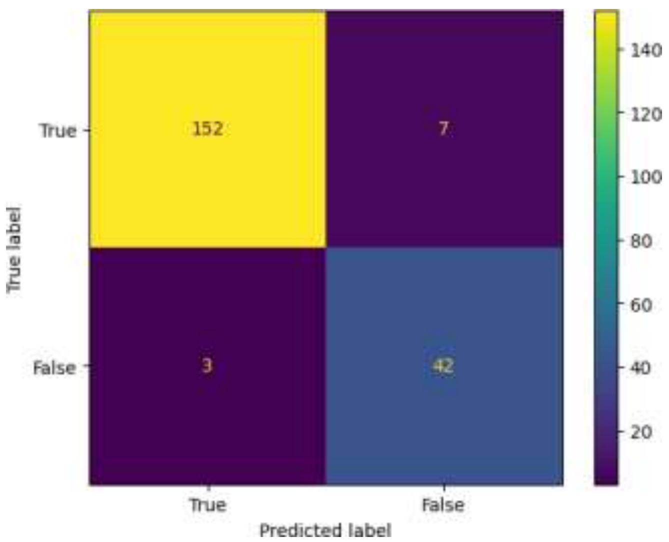


Figure 4.3. Confusion Matrix of KNN

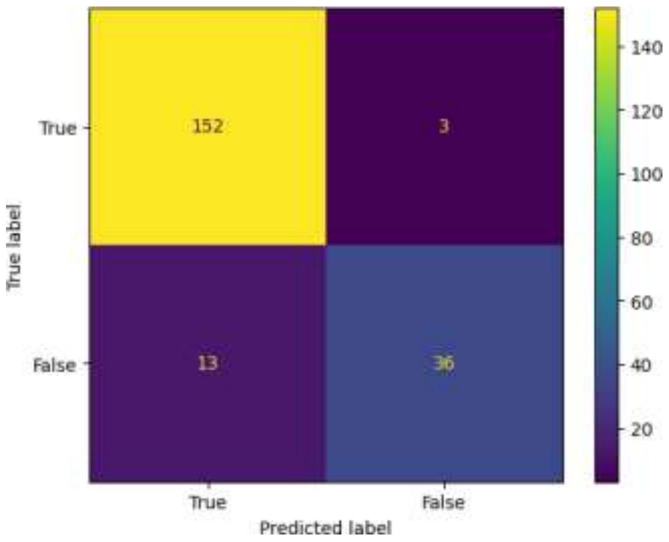


Figure 4.4. Confusion Matrix of Naïve Bayes

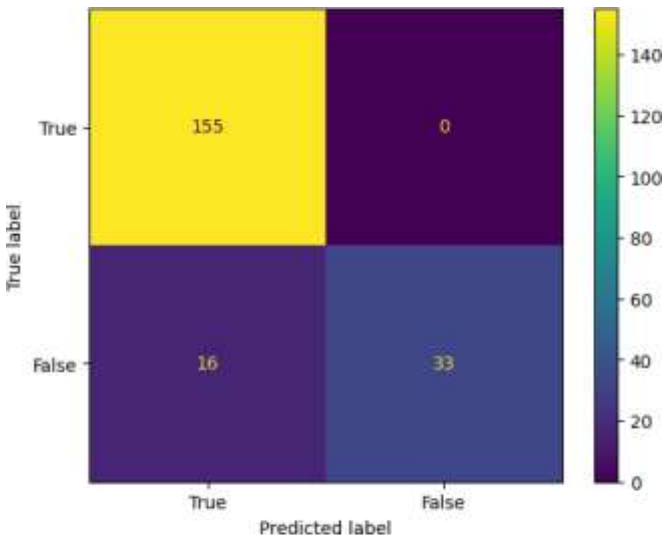


Figure 4.5. Confusion Matrix of SVM

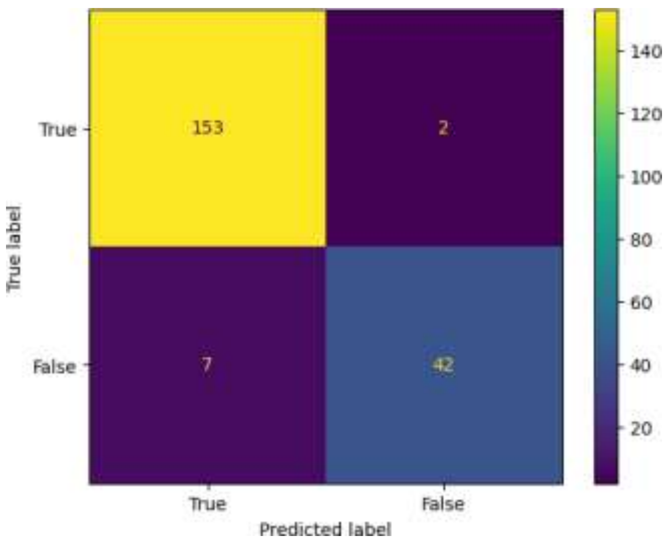


Figure 4.6. Confusion Matrix of XGBoost

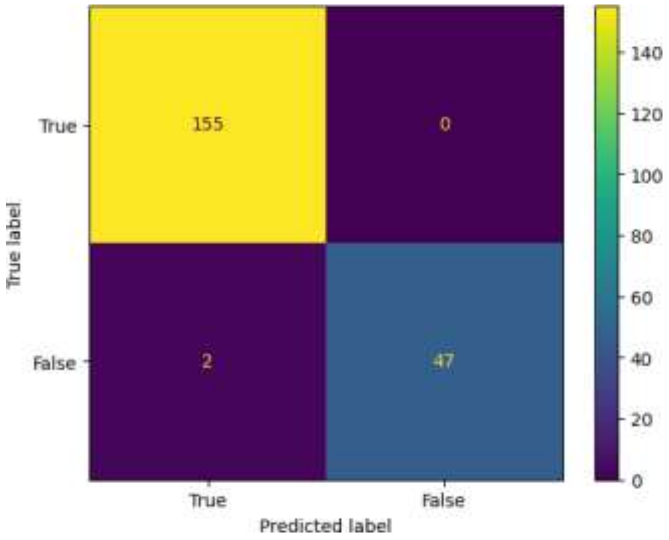


Figure 4.7. Confusion Matrix of Gradient Boosting

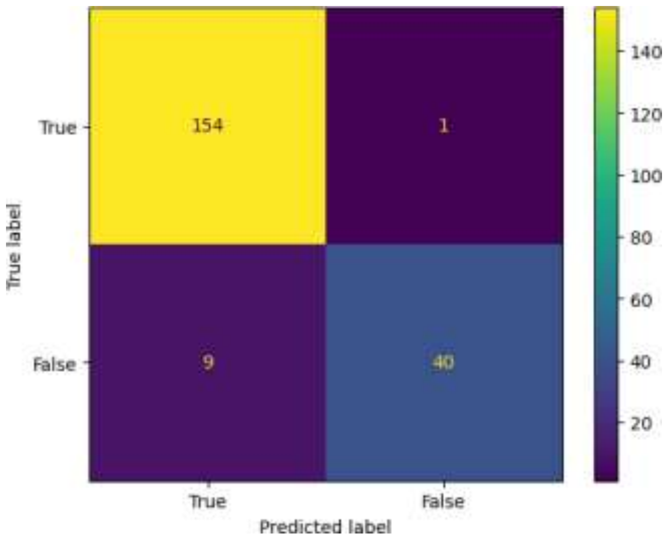


Figure 4.8. Confusion Matrix of Cat Boost

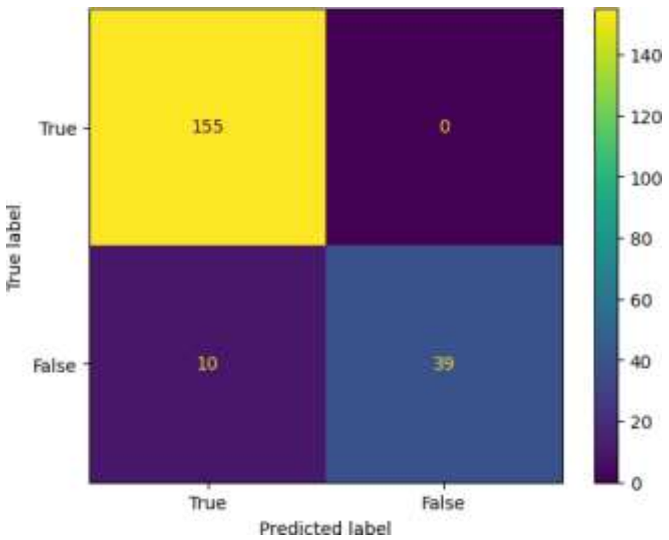


Figure 4.9. Confusion Matrix of Tree Ensemble

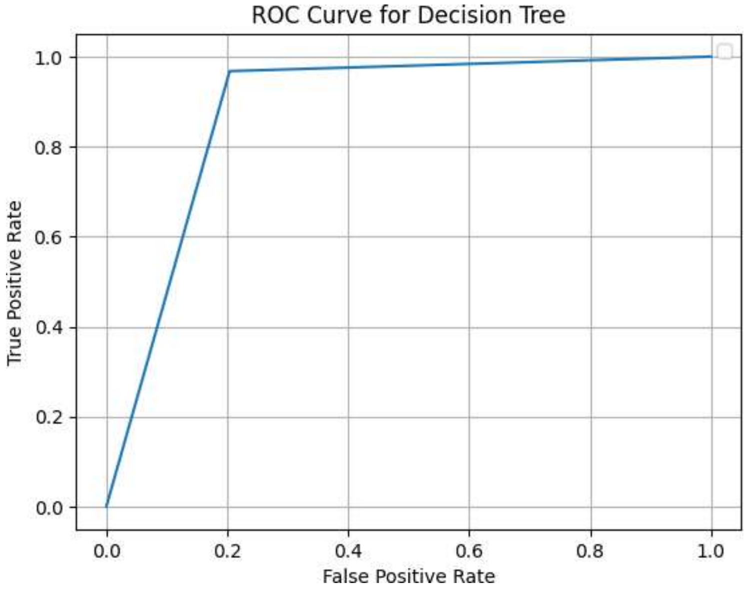


Figure 4.10. ROC Curve of Decision Tree

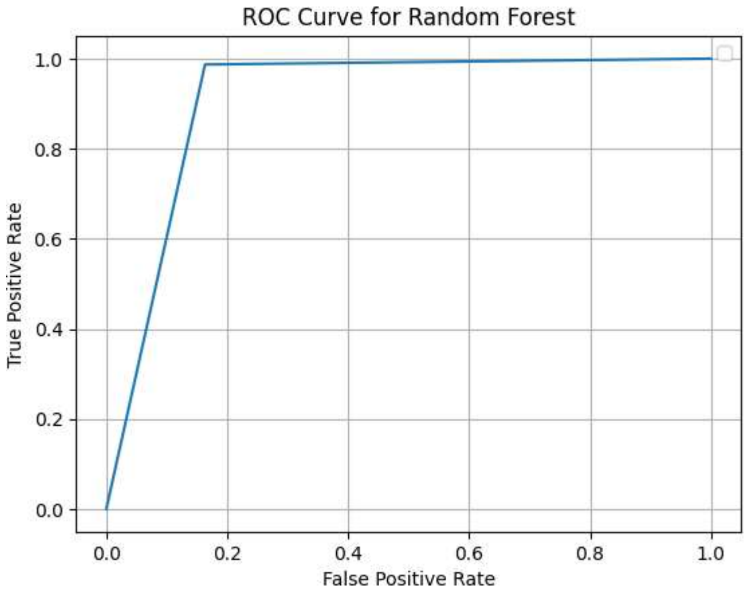


Figure 4.11. ROC Curve of Random Forest

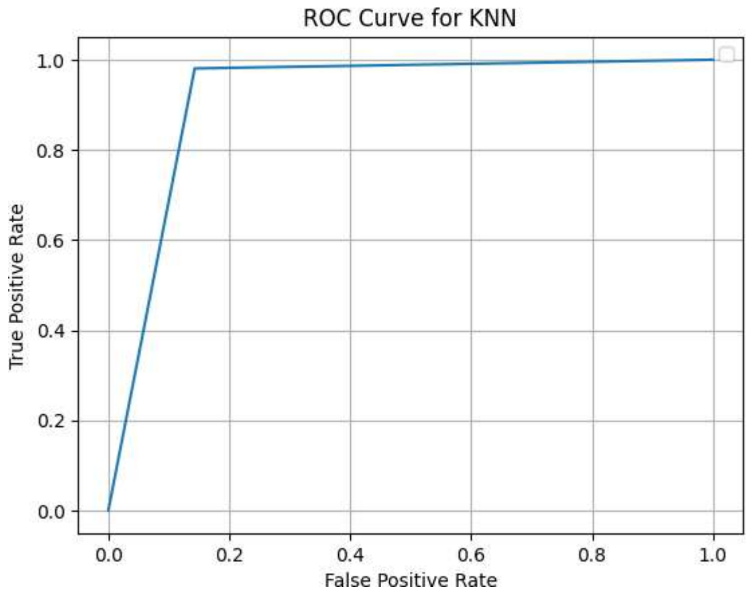


Figure 4.12. ROC Curve of KNN

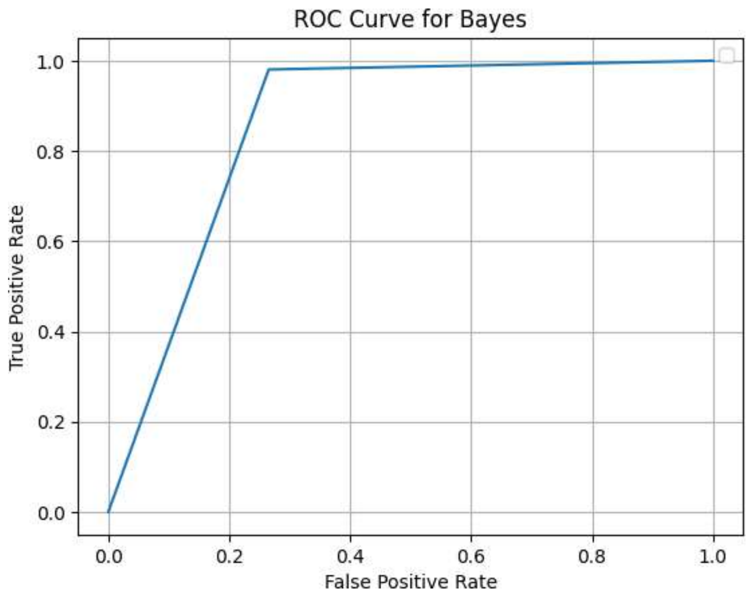


Figure 4.13. ROC Curve of Naïve Bayes

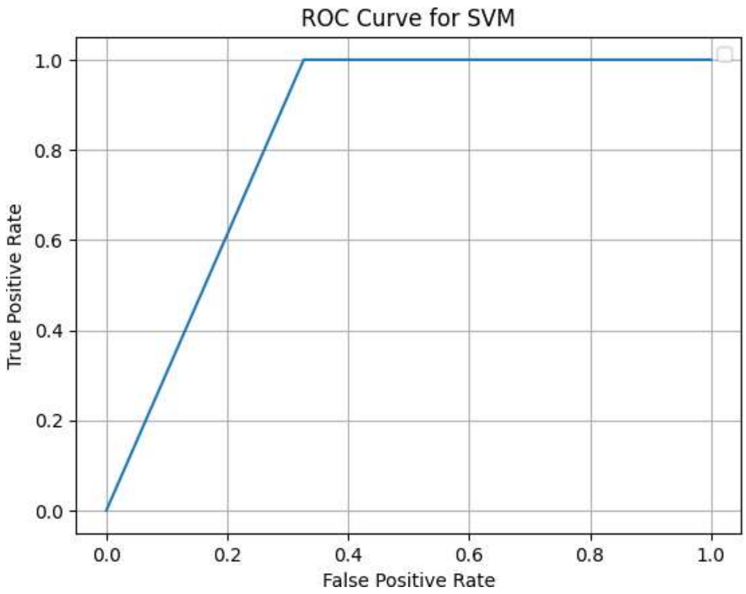


Figure 4.14. ROC Curve of SVM

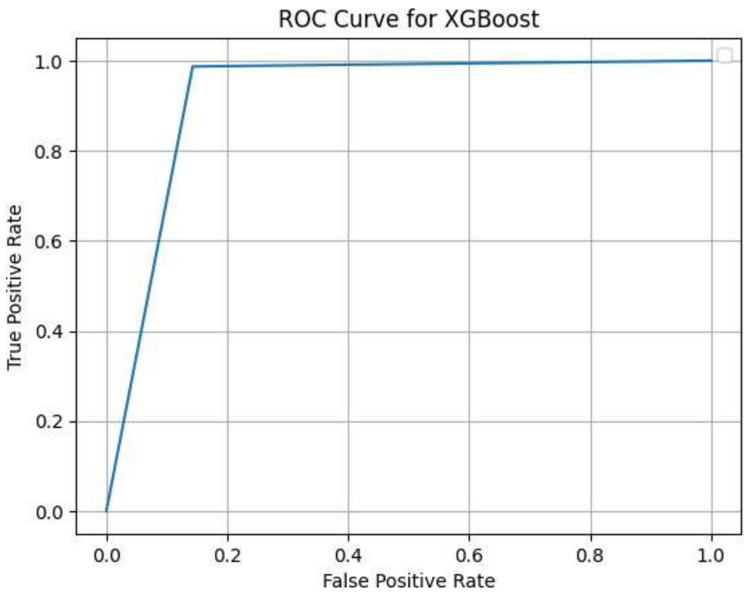


Figure 4.15. ROC Curve of XGBoost

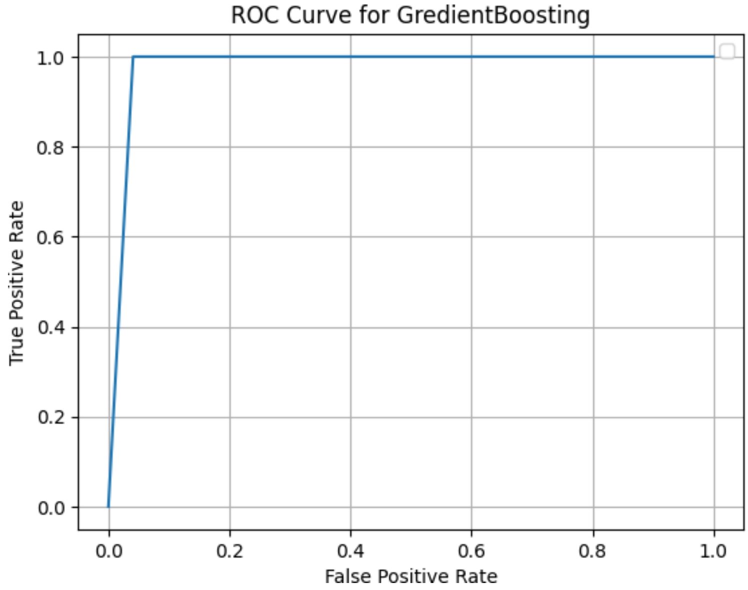


Figure 4.16. ROC Curve of Gradient Boosting

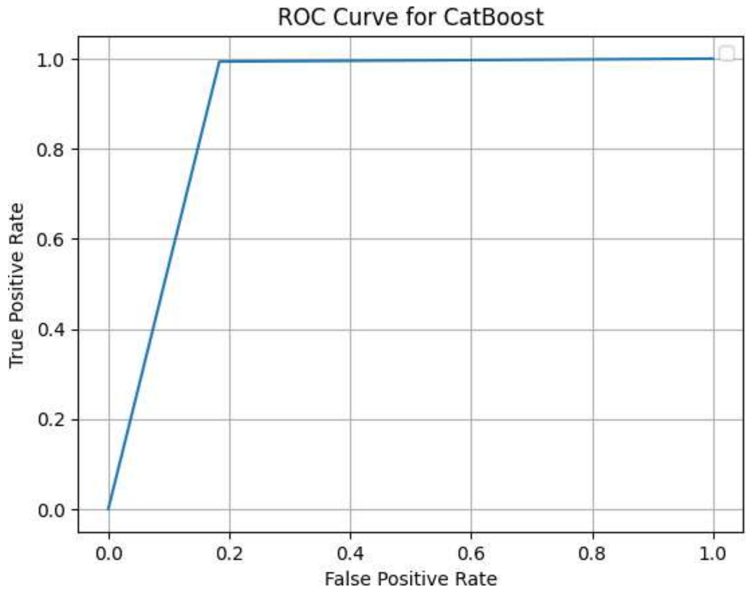


Figure 4.17. ROC Curve of Cat Boost

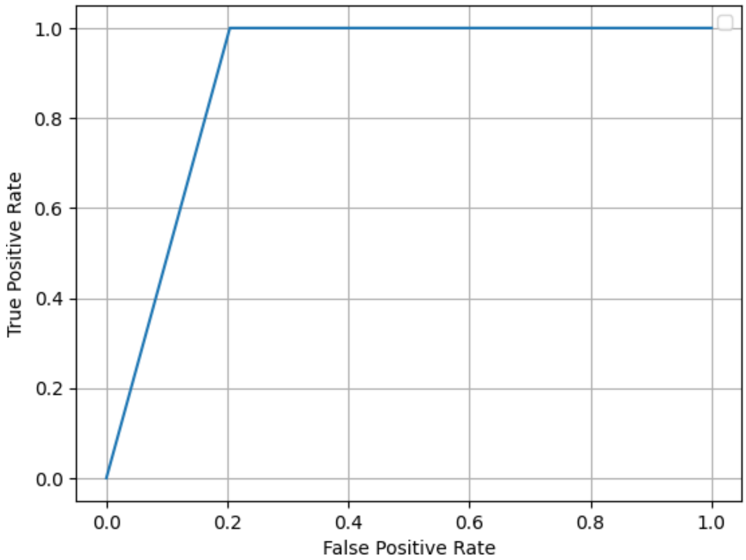


Figure 4.18. ROC Curve of Tree Ensemble

4.3 Results and Discussion

Based on Table 2 & 3, the analysis of machine learning algorithms for insomnia disease detection reveals important insights into their performance. Gradient Boosting is the most consistent and high-performing model across both training and testing metrics, with 100% training score and the highest testing score of 99.01%, paired with excellent precision, recall, and F1 scores (99% across the board). This makes it the optimal choice for this task. XGBoost also performs exceptionally well, with a 100% training score and a testing score of 95.588%, along with 96% precision and 94% F1 score, showing strong generalization ability, though it slightly trails Gradient Boosting in testing performance. Random Forest, KNN, and CatBoost show strong performance with 95% accuracy and 93% F1 scores, though KNN and CatBoost have lower training scores compared to Random Forest and XGBoost. Decision Tree and Naïve Bayes have lower generalization capabilities, with testing scores of 92.65% and 92.15%, respectively, and weaker recall values, which indicate possible overfitting or underperformance on unseen data. SVM struggles with lower recall, highlighting its weaker ability to detect positive instances of insomnia. Overall, Gradient Boosting emerges as the most reliable model, offering a superior combination of accuracy, precision, recall, and F1 score, while XGBoost and Random Forest also provide strong alternatives with slightly different trade-offs.

4.4 Summary

The results and discussion segment of "Insomnia Disease Detection using Machine Learning" highlights the system's performance and its implications for effective insomnia diagnosis. The implemented machine learning models, such as Random Forest, Support Vector Machine, and Neural Networks, demonstrated promising accuracy, precision, recall, and F1 scores in detecting insomnia based on the analyzed physiological and behavioral data. Comparative analysis showed that advanced algorithms like Neural Networks outperformed traditional models in handling complex patterns and providing reliable predictions. Key findings also emphasize the importance of feature selection, such as sleep duration, disruptions, and stress levels, in enhancing model accuracy. The discussion addresses challenges like handling imbalanced datasets and ensuring model interpretability, suggesting potential improvements through techniques like synthetic data generation and explainable AI. The results validate the feasibility of using machine learning for early and accessible insomnia detection, paving the way for scalable solutions in healthcare.

Chapter 5

Engineering Standards and DesignChallenges

5.1 Compliance with the Standards

5.1.1 Software Standards

The required software for this project:

- Google Chrome or Microsoft Edge
- Python 3.9
- Tensor flow
- Jupiter or Google Colab

5.1.2 Hardware Standards

The required software for this project:

- Windows 10 operating system
- Hard Disk 512 GB
- 4 GB RAM

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

The use of machine learning for insomnia disease detection can profoundly impact individuals' lives by improving their overall health, well-being, and productivity. Early and accurate detection of insomnia enables timely intervention, helping individuals manage their symptoms and prevent long-term complications such as chronic fatigue, mental health disorders, and cardiovascular diseases. Accessible diagnostic tools integrated into wearable devices or mobile applications empower individuals to monitor their sleep patterns and take control of their health. By alleviating the burden of insomnia, people can experience better sleep quality, enhanced cognitive function, and increased emotional stability, leading to a more fulfilling and balanced life. Additionally, this technology can reduce the stigma around seeking help for sleep disorders by offering private and user-friendly solutions. Ultimately, machine learning-based insomnia detection has the potential to transform lives by fostering healthier habits, improving personal and professional performance, and enhancing the overall quality of life.

5.2.2 Impact on Society & Environment

5.2.2.1 Impact on Society

The implementation of machine learning for insomnia detection can revolutionize public health by addressing the widespread challenges associated with this sleep disorder. By providing early and accurate diagnosis, these systems can reduce the prevalence of associated conditions such as depression, anxiety, and cardiovascular diseases, fostering a healthier population. The integration of such tools into wearable devices and mobile applications increases accessibility, particularly for underserved communities, thereby reducing healthcare disparities. Moreover, the economic impact is significant—early detection and intervention can alleviate productivity loss in workplaces caused by sleep deprivation, benefiting both individuals and economies. Additionally, the availability of affordable and automated diagnostic tools can ease the burden on healthcare systems, enabling them to focus resources on other critical areas, ultimately leading to a more resilient and equitable society.

5.2.2.2 Impact on the Environment

The development and deployment of machine learning-based systems for insomnia detection can also have environmental implications. By minimizing the need for traditional, resource-intensive diagnostic methods such as polysomnography, these tools reduce the energy consumption and material waste associated with sleep laboratories. Furthermore, the shift towards digital and remote solutions decreases the need for travel to medical facilities, reducing carbon emissions and supporting sustainability. However, the environmental impact of producing wearable devices and the energy consumption of data centers powering machine learning algorithms must also be addressed. By prioritizing energy-efficient computing and sustainable production practices, this technology can contribute positively to both healthcare and environmental sustainability, creating a balanced approach to innovation.

5.2.3 Ethical Aspects

The use of machine learning for insomnia detection raises several ethical considerations that must be addressed to ensure responsible and equitable implementation. Data privacy and security are paramount, requiring robust measures to protect sensitive personal information and compliance with regulations like GDPR and HIPAA. Informed consent is crucial, ensuring users are aware of how their data will be used and shared. Bias and fairness must be mitigated to prevent unequal outcomes across diverse populations, while accuracy and reliability are essential to avoid harm from incorrect diagnoses. Accessibility and equity should be prioritized to ensure the technology benefits all, including underserved communities, without being cost-prohibitive. Clear accountability frameworks are necessary to handle system errors, and transparency in the decision-making processes of machine learning models fosters trust among users and healthcare providers. Additionally, ethical use of the technology must be upheld, ensuring it is solely aimed at improving health outcomes and not exploited for commercial purposes without user consent. Addressing these aspects ensures the technology aligns with principles of fairness, transparency, and societal benefit.

5.2.4 Sustainability Plan

A sustainability plan for insomnia disease detection using machine learning focuses on creating a system that is economically, socially, and environmentally viable over the long term. Economically, the system should be cost-effective, leveraging scalable cloud-based platforms and open-source technologies to minimize development and maintenance expenses. Socially, it must prioritize accessibility and inclusivity, ensuring underserved populations have equal access to its benefits through affordable or subsidized solutions. Environmentally, the plan should emphasize energy-efficient algorithms, sustainable production practices for wearable devices, and the use of renewable energy in data centers to minimize the carbon footprint. Regular updates and retraining of the machine learning models with diverse, anonymized datasets will ensure continued accuracy, fairness, and relevance. Partnerships with public health organizations, NGOs, and governments can provide ongoing support and funding, while education and awareness campaigns will foster user trust and adoption. By integrating these elements, the sustainability plan ensures the long-term effectiveness, scalability, and positive impact of the system.

5.3 Project Management and Financial Analysis

- Project Objectives:
    A. To develop a web-based application for verify vehicle damage
    B. To make short & easy the insurance claims process
- Project Timeline:
    A. Phase 1: Project Planning and Research (1-2 weeks)
    B. Phase 2: Established Collaboration with Professionals (4-5 weeks)
    C. Phase 3: Reference Paper Collection (4-6 weeks)
    D. Phase 4: Paper Review (4-6 weeks)
    E. Phase 5: Data Collection (8-10 weeks)
    F. Phase 6: Data Analysis (4-6 weeks)
    G. Phase 7: Data Preprocessing (3-4 weeks)
    H. Phase 8: Model Implement (3-4 weeks)
    I. Phase 9: Model Evaluation (On Going)

J. Phase 10: Prototype Design (3-4 weeks)
K. Phase 11: Front End Development (On Going)
L. Phase 12: Back End Development (Up Coming)
M. Phase 12: Deployment & Testing (Up Coming)
N. Phase 14: Post-Launch & Marketing (Up Coming)

- Resource Planning:

A. Equipment and Tools:

- Development and Testing Servers
- High-performance Computers for Development Team Design Software (e.g., Adobe Creative Suite)
- Collaboration Tools (e.g., Slack, Trello, or project management software)
- Version Control System (e.g., Git)
- Testing Tools (e.g., Selenium for automated testing)

B. Software and Technologies:

- Front-End Technologies (e.g., HTML, CSS, JavaScript, React or Angular)
- Back-End Technologies (e.g., Node.js, Django, Flask, or Ruby on Rails)
- Database Management System (e.g., MySQL, PostgreSQL, or MongoDB)
- Server Hosting (e.g., AWS, Azure, or Google Cloud)
- Security Software and SSL Certificates

C. Data and Content:

- Product Images: Obtained through agreements with suppliers
- User Documentation: Prepared by the technical writing team

D. Training and Skill Development:

- Ensure that the development team has the necessary training and skills in web-based application development, security, and database management.
- Provide additional training on specific technologies and tools as needed.

● Communication Plan:

A. Stakeholder Meetings:

Purpose: Update stakeholders on project requirements gathering progress and gather feedback.

Participants: Team members, and stakeholders.

Frequency: Bi-weekly or as specified in the project plan.

B. Change Control Meetings:

Purpose: Discuss and approve any changes to the project scope or requirements.

Participants: Supervisor and team members.

Frequency: As needed when change requests arise.

Finance: The cost table is given below:

Table 5.1: Cost Estimated Table

SN	Components	Estimated Cost (BDT)
1	Visiting Stakeholders	2500-3000
2	Software and Tools	5000-7000
3	Data Collection and Processing	2500-3000
4	Documentation and Report Writing	1500-2000
5	Contingency (10% of total)	1000- 1500
Total Estimated Cost		12500-16500

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

Table 5.2: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of ApplicableCodes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
✓	✓	✓	✓			✓

This project demonstrates EP1 by achieve K3, K4, K5, K6 & K8.

This project addresses EP2 by recognizing the hurdles in insomnia detection, including limitations of traditional methods and the complexities of integrating machine learning. Through comparative analysis, it confronts challenges in understanding spatial distributions, offering insights for refining diagnostic methodologies.

This project addresses EP3 by meticulously comparing experimental outcomes, highlighting Machine Learning as the chosen significant solution for enhancing Insomnia detection amidst multiple potential approaches.

This project's interdisciplinary approach extends beyond computer science and engineering, impacting medical diagnostics in respiratory diseases like insomnia, contributing to advancements in public health and healthcare practices which indicates EP4.

This project's comprehensive approach addresses high-level problems by integrating various components across data collection, statistical analysis, and proposed methodology, ensuring a holistic solution to complex challenges in medical diagnostics which ensures EP7.

Mapping with Knowledge Profile for EP1

Table 5.3: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
✓	✓	✓	✓	✓

This project demonstrates fundamental engineering (K3) principles by employing machine learning models, data preprocessing for classification task. The project demonstrates specialist knowledge (K4) by conducting boosting models, also ensemble models, enhancing insomnia detection accuracy, crucial for computer-aided diagnosis.

The project applies engineering practice & design (K5) by the figure of process of experiments. The project addresses engineering practice & technology (K6) by employing ML models.

This project ensures to K8 (Research Literature) by synthesizing insights from recent studies, to advance insomnia detection using deep learning, showcasing a comprehensive understanding of current methodologies.

5.4.2 Engineering Activities

Table 5.4: Mapping with complex engineering activities.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓			✓	✓

Our project utilizes diverse resources such as high-performance computing infrastructure, GPUs, deep learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements in insomnia detection through machine learning.

This project contributes to society by improving healthcare through advanced insomnia detection methods, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for patient data privacy.

This project expands upon existing research by examining a novel approach in insomnia detection through machine learning, demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field.

5.5 Summary

Chapter 5 discusses the engineering standards and challenges associated with the design and implementation of the "Insomnia Disease Detection using Machine Learning" system. It outlines the compliance with software and hardware standards necessary for the project's success, including the use of Python 3.9, TensorFlow, and computing systems with adequate processing capabilities. This chapter emphasizes the societal, environmental, and ethical impacts of the project. On a societal level, the integration of machine learning into healthcare promotes accessibility, equity, and improved diagnostic accuracy. From an environmental perspective, the system minimizes the reliance on resource-intensive diagnostic methods, although considerations around energy-efficient computation and the environmental impact of wearable devices are acknowledged.

The chapter also addresses ethical concerns, such as data privacy, fairness, and accessibility, highlighting the importance of informed consent and transparent algorithms. Sustainability is a key focus, with a plan to ensure the project's economic, social, and environmental viability. This includes leveraging open-source tools, prioritizing energy-efficient algorithms, and fostering inclusivity. Challenges encountered during implementation, such as handling complex datasets and ensuring model generalization, are detailed, along with a summary of the solutions employed. By adhering to these standards and addressing challenges, the project demonstrates its commitment to creating a reliable, ethical, and impactful solution for insomnia detection using machine learning.

Chapter 6

Conclusion

6.1 Summary

This study demonstrates the effectiveness of various machine learning algorithms in detecting insomnia disease. Among the models evaluated, Gradient Boosting stands out as the most reliable and robust, achieving the highest performance across both training and testing metrics, with a 99.01% testing score, 99% precision, and 99% F1 score. This indicates its exceptional ability to generalize well to unseen data while maintaining high accuracy in predicting insomnia cases. XGBoost, Random Forest, and KNN also show strong results, offering a good balance of precision, recall, and F1 scores, making them viable alternatives depending on specific model preferences. On the other hand, Naïve Bayes and SVM show lower performance, particularly in recall, highlighting challenges in detecting positive cases of insomnia. Overall, Gradient Boosting is the recommended model for insomnia detection due to its superior balance of accuracy, precision, and recall, but other models like XGBoost and Random Forest may also be considered based on their performance and specific use case requirements. Further research and fine-tuning of these models may yield even better results, but the findings of this study provide a solid foundation for selecting appropriate machine learning techniques in healthcare applications for insomnia detection.

6.2 Limitation

The implementation of the "Insomnia Disease Detection using Machine Learning" system faces several limitations that warrant consideration. The dataset used for model training and testing may lack diversity, potentially hindering the system's ability to generalize across different populations, such as varied age groups, cultural backgrounds, and health conditions. Furthermore, imbalanced datasets, with a disproportionate representation of insomniac and non-insomniac cases, could bias the model’s predictions, particularly in underrepresented scenarios. The interpretability of some advanced machine learning models, such as deep learning algorithms, poses challenges, making it difficult for healthcare professionals to trust and understand the system’s decision-making process. Additionally, the reliance on specific physiological and behavioral features may restrict the system's applicability if these data points are unavailable or inaccurate in real-world applications. Ethical and privacy concerns related to handling sensitive personal data also present barriers to widespread adoption, necessitating robust data protection measures and compliance with legal frameworks.

6.3 Future Work

To overcome these limitations, future work should focus on expanding the dataset to include diverse populations, ensuring the system’s broad applicability and reliability. Advanced techniques like explainable AI and federated learning could improve model interpretability and privacy, fostering trust and compliance with ethical standards. Real-time applications integrated into wearable devices and mobile platforms would enhance accessibility and usability, enabling continuous monitoring and timely interventions. Incorporating multi-modal data, such as environmental factors and genetic predispositions, could further increase the robustness and accuracy of the detection system. Collaborations with healthcare professionals for clinical validation and deployment in real-world settings will ensure the system’s practicality and effectiveness. Lastly, developing user feedback loops and adaptive mechanisms would refine predictions and personalize the system for individual users, making it a comprehensive and impactful tool for managing insomnia.

References

[1] *Dhakatribune.com*. [Online]. Available: https://www.dhakatribune.com/bangladesh/295481/two-thirds-of-university-students-suffer-from. [Accessed: 12-Dec-2024].

[2] A. Calderon *et al.*, “Machine learning and Bayesian network analyses identifies associations with insomnia in a national sample of 31,285 treatment-seeking college students,” *BMC Psychiatry*, vol. 24, no. 1, p. 656, 2024.

[3] M. Ingle *et al.*, “A systematic review on automatic identification of insomnia,” *Physiol. Meas.*, vol. 45, no. 3, p. 03TR01, 2024.

[4] M. M. Islam, A. K. M. Masum, S. Abujar, and S. A. Hossain, ―Prediction of chronic Insomnia using Machine Learning Techniques,‖ in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2020.

[5] C. Jansen, T. Penzel, S. Hodel, S. Breuer, M. Spott, and D. Krefting, ―Network physiology in insomnia patients: Assessment of relevant changes in network topology with interpretable machine learning models,‖ *Chaos*, vol. 29, no. 12, p. 123129, 2019.

[6] Y. Furukawa *et al.*, ―Cognitive behavioral therapy for insomnia to treat major depressive disorder with comorbid insomnia: A systematic review and meta-analysis,‖ *J. Affect. Disord.*, vol. 367, pp. 359–366, 2024.

[7] C. Baglioni *et al.*, ―Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies,‖ *J. Affect. Disord.*, vol. 135, no. 1–3, pp. 10–19, 2011.

[8] R. Alazaidah, G. Samara, M. Aljaidi, M. Haj Qasem, A. Alsarhan, and M. Alshammari, ―Potential of machine learning for predicting sleep disorders: Acomprehensive analysis of regression and classification models,‖ *Diagnostics (Basel)*, vol. 14, no. 1, 2023.

[9] H. Doos Ali Vand, B. Gharraee, A.-A. Asgharnejad Farid, and M. Ghaleh Bandi, ―Prediction of insomnia severity based on cognitive, metacognitive, and emotional variables in college students,‖ *Explore (NY)*, vol. 10, no. 4, pp. 233–240, 2014.

[10] M. Shahin, L. Mulaffer, T. Penzel, and B. Ahmed, ―A two stage approach for the automatic detection of insomnia,‖ *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2018, pp. 466–469, 2018.

[11] S. Li, P. Zhu, G. Cai, J. Li, T. Huang, and W. Tang, ―Application of machine learning models in predicting insomnia severity: an integrative approach with constitution of traditional Chinese medicine,‖ *Front. Med. (Lausanne)*, vol. 10, p. 1292761, 2023.

[12] Q. Lu *et al.*, ―Connectomic disturbances underlying insomnia disorder and

predictors of treatment response,‖ *Front. Hum. Neurosci.*, vol. 16, p. 960350, 2022.

[13] T. Roth, —Insomnia: definition, prevalence, etiology, and consequences,‖ *J. Clin. Sleep Med.*, vol. 3, no. 5 Suppl, pp. S7-10, 2007.

[14] C. H. Bastien, A. Vallières, and C. M. Morin, —Validation of the Insomnia Severity Index as an outcome measure for insomnia research,‖ *Sleep Med.*, vol. 2, no. 4, pp. 297–307, 2001.

[15] R. B. Berry *et al.*, —Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine: Deliberations of the sleep apnea definitions task force of the American academy of sleep medicine,‖ *J. Clin. Sleep Med.*, vol. 8, no. 5, pp. 597–619, 2012.

[16] S. M. Mckinney, —Machine learning in healthcare: potential and challenges,‖ *Journal of Sleep Research*, vol. 29, no. 4, 2020.

[17] T. Kuo, —Application of machine learning algorithms for the detection of insomnia based on physiological signals and sleep architecture,‖ *Journal of Biomedical Informatics*, vol. 102, 2020.

[18] A. Supratak, H. Dong, C. Wu, and Y. Guo, —DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG,‖ *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1998–2008, 2017.

Insomnia Disease Detection using Machine Learning

ORIGINALITY REPORT

12%	7%	6%	9%
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	Submitted to University of Finance – Marketing Student Paper	2%
4	V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challenges in Information, Communication and Computing Technology", CRC Press, 2024 Publication	1%
5	Submitted to Higher Education Commission Pakistan Student Paper	1%
6	v1.overleaf.com Internet Source	1%
7	Manisha Ingle, Manish Sharma, Kamlesh Kumar, Prince Kumar, Ankit Bhurane, Heather Elphick, Deepak Joshi, U Rajendra Acharya. "A	1%

systematic review on automatic identification of insomnia", Physiological Measurement, 2024
Publication

8	Sheetanshu Gupta, Dhirendra Kumar, Radhika Negi, Ranjan Singh, Mohammad Javed Ansari, Shakuli Kashyap, Sudhir Mehrotra. "Genomic Intelligence: Metagenomics and Artificial Intelligence", CRC Press, 2024 Publication	1%
9	tnsroindia.org.in Internet Source	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	Off		