Predicting Sleep Disorders Using a Stacking Ensemble of Gradient Boosting, Decision Tree, and Logistic Regression Models

Submitted By
Md. Jahidul Islam joy
ID: 213-35-800
Department of Software Engineering,
Daffodil International University.

Supervised By
Afsana Begum
Assistant Professor & Co-Ordinator of M.Sc
Department of Software Engineering,
Daffodil International University.

This report is presented in partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering.

Summer – 2025

APPROVAL

This thesis titled on "**Predicting Sleep Disorders Using a Stacking Ensemble of Gradient Boosting, Decision Tree, and Logistic Regression Models**", submitted by **Student Name (ID: 213-35-800)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

______________________________ **Chairman**

Dr. Imran Mahmud
Professor & Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

______________________________ **Internal Examiner 1**

Md Shohel Arman
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

______________________________ **Internal Examiner 2**

Md. Rajib Mia
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

______________________________ **External Examiner**

Md Habibur Rahman
Associate Professor
Department of Computer Science and Engineering
Islamic University, Bangladesh

DECLARATION

I hereby declare that this thesis report is done by me under the supervision of Afsana Begum, Assistant Professor, Department of Software Engineering, Daffodil International University, in fulfillment of my original work. I am also declaring that, to the best of my knowledge, neither this thesis nor any part thereof has been submitted elsewhere for the award of an M.Sc. or any degree.

Supervised by:

15.09.25

Assistant Professor & Co-Ordinator of M.Sc
Department of SWE
Daffodil International University

Submitted by:

Joy 15.09.25

Md. Jahidul Islam Joy
ID: 213-35-800
Batch: 36th
Department of SWE
Daffodil International University

ACKNOWLEDGEMENT

I extend my heartfelt gratitude to my research supervisor, **Mst. Afsana Begum**, for her invaluable guidance, insightful feedback, and unwavering support throughout the entire research process. Her encouragement and expertise have been instrumental in shaping this work.

I would also like to convey my sincere appreciation to **Dr. Imran Mahmud**, Head of the Faculty, for providing the necessary facilities and a supportive academic environment to carry out this research. My thanks also go to all the **lecturers** of the Department of Software Engineering for their valuable advice and kind assistance whenever I encountered challenges.

Lastly, I wish to express my deepest respect and gratitude to my beloved **parents** for their constant love, sacrifices, and moral support, without which this achievement would not have been possible.

Abstract

sleep disorders, including insomnia, obstructive sleep apnea, and restless leg syndrome, have significant impacts on human health and productivity. Conventional diagnostic methods, such as polysomnography, are resource-intensive and not suitable for large-scale screening. Recent advancements in machine learning (ML) offer promising alternatives by enabling automated, data-driven diagnosis from lifestyle and health-related parameters. This study presents a comprehensive investigation of existing ML-based approaches for sleep disorder detection, followed by the development of an optimized diagnostic framework. The proposed model integrates advanced preprocessing—feature selection, class balancing via SMOTE-ENN, and normalization—with a Stacking Classifier combining Gradient Boosting, Decision Tree, and Logistic Regression to enhance predictive performance. Evaluation on the Sleep Health and Lifestyle dataset achieved an accuracy of 98.67%, precision of 98.80%, recall of 98.60%, and F1-score of 98.70%, outperforming baseline classifiers and previous literature benchmarks. The framework fills research gaps by making models clearer and easier. It also helps reduce the effect of class imbalance in the data. The design keeps the tool low cost and very easy to use. The system works well with larger data, so it can grow fast. It also helps with the early detection of health problems. The method can support preventive care and also improve patient safety. In this way, the framework is useful for both experts and normal users. It gives support for real healthcare needs and makes solutions simpler.

Keywords — Sleep Disorders, Machine Learning, Stacking Classifier, SMOTE-ENN, Sleep Health and Lifestyle Dataset, Classification, Data Preprocessing, Healthcare Analytics

Table Of Contents

Tables

Figures

Chapter 1

1. Introduction

Sleep is a biological necessity that underpins physical recovery, cognitive performance, and emotional stability. Disruptions in sleep patterns can lower daily performance and harm long-term health (Berry et al., 2020). Such disruptions are linked with hypertension, diabetes, heart problems, and mental disorders. Modern lifestyles are often stressful and inactive, which increases the risk of sleep disorders. Because of this, detecting sleep problems early and accurately has become very important.

Traditional methods like polysomnography (PSG) are used to diagnose sleep problems. However, these methods are slow, costly, and need many resources. They also need expert help and special clinical settings, which limits easy access. Some statistical and computer methods try to support these diagnoses, but they often fail. Their prediction quality becomes weak when used on large and diverse medical data.

Recent growth in artificial intelligence and machine learning brings new hope for solving these issues. Machine learning can find hidden patterns in big data and support clinical judgment. However, single models like logistic regression, decision trees, and support vector machines give unstable results (Tareq, 2024; Nuraeni & Faisal, 2025). They often cannot handle complex and high-dimensional health data well. There is a strong need for methods that mix different models and improve accuracy.

This study builds a stacking ensemble model using Gradient Boosting, Decision Tree, and Logistic Regression. The aim is to use the strengths of each model and reduce their limits. The model is tested on organized sleep health datasets to prove its strength. These methods can improve predictions, support clinical use, and advance healthcare analytics.

1.2 Background and Motivation

Sleep Disorders and Their Global Burden

The sleep disorders of insomnia, obstructive sleep apnea, restless leg syndrome, and hypersomnia are common in most individuals all over the world. These health complications may predispose one to metabolic syndrome and heart diseases. They are also associated with mental disorders that have the potential to interfere with emotional well being and behavior. Undiagnosed or untreated sleeping issues usually reduce the productivity of the work and increase the risks of accidents (Berry et al., 2020). These issues are also likely to lower the quality of life and impose excessive social and financial burdens. The increasing sensitivity to such impacts indicates a necessity of proper prediction instruments. Sleep issues can be discovered and treated at an earlier stage with the help of reliable methods.

Limitations of Conventional Approaches

The use of traditional diagnosis relying on the clinical observation of the patient, survey, and testing of polysomnography is usual. These procedures are effective in the controlled environment but face severe difficulties. The major barriers include high costs, reliance on special machines and difficulty of access. There is also heavy subjectivity in how the results are understood by different experts. These issues leave many people undiagnosed and without needed medical support or treatment. Even diagnosed patients may not get quick help due to slow and complex procedures. The slow process can make health risks worse and delay needed care.

Motivation for Machine Learning-Based Approaches

Machine learning has become useful in healthcare because it can handle large datasets. It can detect patterns and make predictions without requiring strict manual rules or coding. In sleep health, such methods can analyze clinical, demographic, and behavior data to classify risk groups. Machine learning can scale diagnoses for more people and reduce the workload of doctors. However, most current models are made for small and single datasets (Airlangga, 2024; Alshammari, 2024). These models often give inconsistent results and fail when tested on new patient data. Such limits show the need for stronger and broader approaches to improve predictions.

Motivation for Ensemble Learning

Ensemble learning mixes several models to lower bias, variance, and prediction errors. It uses the strengths of different models and gives more stable and accurate outcomes. Methods like bagging, boosting, and stacking are used in many fields but rarely in sleep studies. Stacking is very useful because it lets models support one another using a meta-model. It can improve predictions by combining outputs from base models to make better final decisions.

1.3 Problem Statement

Sleep disorders affect millions of people around the world and remain hard to detect. Conventional methods like polysomnography and manual scoring are still the gold standard worldwide. However, these methods are slow, expensive, and rely on human experts for interpretation (Berry et al., 2020; ICAT, 2023). Human errors and time-consuming steps often lead to inconsistent and incorrect diagnoses. Researchers have tried automated diagnostic tools that use single models such as logistic regression or decision trees. These models fail to find complex nonlinear patterns in sleep data and give low accuracy (Tareq, 2024; Hidayat, 2023; Nuraeni & Faisal, 2025).

Ensemble learning is a powerful solution for improving prediction accuracy and reducing errors. It has shown great results in many fields but is rarely used in sleep research. Past studies explored random forests, support vector machines, and optimization-driven models (Alshammari, 2024; Alhussan, 2025). Some studies even used voting classifiers to combine predictions (SSRN, 2024). However, only a few have used stacking ensembles that mix Gradient Boosting, Decision Trees, and Logistic Regression (Riyandi et al., 2025). The lack of stacking-based studies shows a clear gap in research on sleep disorder prediction.

Another challenge comes from high false-positive and false-negative rates in current approaches. Such errors can result in wrong treatment plans and pose risks to patient safety (Airlangga, 2024; IEEE Access, 2025). These issues can also reduce trust in automated systems and delay clinical acceptance. In addition, many models fail to generalize well to new populations and bigger datasets. They lose accuracy when used outside their original data (Selvakani et al., 2025; Rahman et al., 2025). Overcoming these problems needs robust ensemble models that combine strengths of many classifiers. Such models can offer accurate, scalable, and clinically meaningful predictions for diverse populations.

Chapter 2

2. Literature Review

2.1: Existing Techniques for Detection of Sleep Disorders

Polysomnography is generally regarded as the gold standard of diagnosing sleep disorders (Berry et al., 2020). Nevertheless, the research points at its weaknesses, in specific cost and access, aspects. The researchers have thus attempted to come up with scalable alternatives in form of computational models that are built on clinical and survey data (ICAT, 2023).

2.2 Machine Learning in the Diagnosis of Sleep Disorders

Including the machine learning models, Random Forests (Tareq, 2024; Hidayat, 2023), Support Vector Machines (Nuraeni and Faisal, 2025) and K-means clustering (Septiadi et al., 2025) have been demonstrated to deliver better predictive accuracy than the traditional models. Optimization-driven models, including metaheuristic methods (Alhussan, 2025), have further enhanced classification results. Yet, reliance on single algorithms often results in overfitting, high variance, or limited interpretability.

2.3 Previous Studies and Research

Recent studies have explored ensemble learning methods. Voting classifiers (SSRN, 2024) and optimized ensembles (Rahman et al., 2025; IEEE Access, 2025) have shown that combining multiple models can outperform individual classifiers. However, most prior work has not fully leveraged stacking approaches. Very few studies have integrated Gradient Boosting, Decision Tree, and Logistic Regression into a unified framework for sleep disorder prediction (Riyandi et al., 2025). This research seeks to address that gap.

2.4 Research Gap

Many research has been done so far by using machine learning algorithms for early detection of Sleeping Disorder. But it may have some causes or issues found at analysis of current paper. Understanding those issues can be very useful for future work & further improvement in this field. Although machine learning has advanced predictive modeling in sleep health, the majority of existing studies employ single-model approaches, which limit accuracy and generalizability. Ensemble learning, particularly stacking, remains underutilized in this field. Therefore, there is a pressing need for a stacking-based methodology that can integrate multiple classifiers to improve diagnostic accuracy, minimize false classifications, and ensure applicability across diverse datasets.

Chapter 3

3. Methodology

Given the number of lifestyle and health factors, this study investigates the application of machine learning algorithms to predict sleep disorders. GradientBoostingClassifier for sleep disorder prediction: We diagnose sleep disorders using the Sleep Health and Lifestyle Dataset and a variety of supervised machine learning models. The main 4 phases are data collection and preprocessing, model selection and training, model evaluation, and hyperparameter tuning for the methodology.

3.1 Methodology Diagram (Possible Method Diagram)

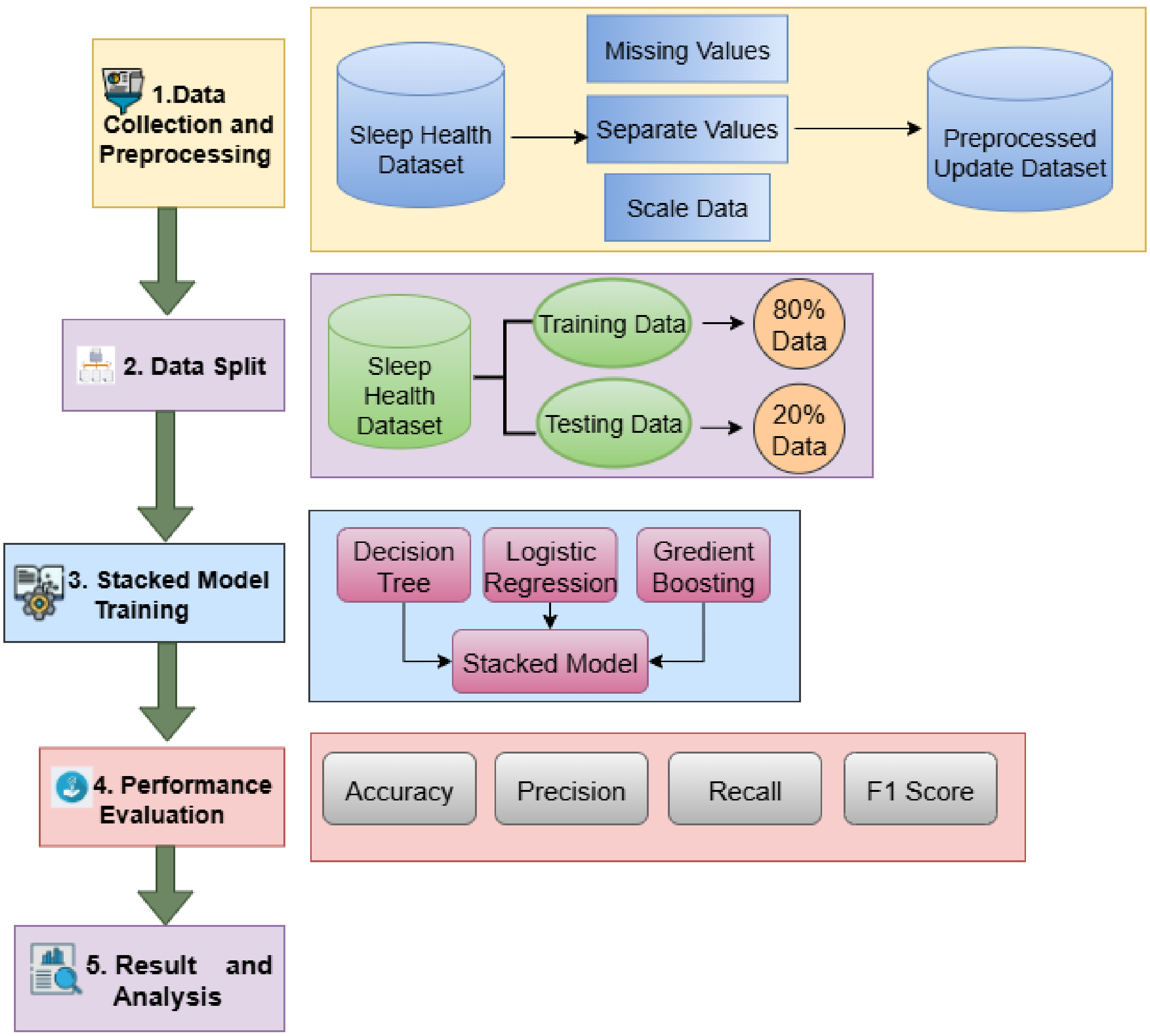


Figure 0: Methodology Workflow

3.2 Data Collection

The study used the Sleep Health and Lifestyle Dataset collected from Kaggle for analysis. The dataset offers detailed information about people's sleep habits and daily lifestyle behaviors. It includes 374 records and 13 columns like age, gender, stress level, sleep duration, and physical activity. The main goal was to predict sleep disorder status with high accuracy. The classification focused on identifying normal sleep, insomnia, and sleep apnea for accurate and reliable detection of conditions.

3.3 Dataset Description

The current research uses the Sleep Health and Lifestyle Dataset containing real sleep health information. It has 374 samples that represent different groups of people from various backgrounds. The dataset includes both numerical and categorical values linked with many sleep health aspects. Such data can help in understanding how lifestyle choices affect different sleep disorder conditions. It can also guide in improving healthy habits that support better focus and educational achievements.

Dataset Features:

- Number of Samples: 374
- Target Variable: Sleep Disorder (Categorical)

Categories:

- None
- Sleep Apnea
- Insomnia

The dataset includes **13** features that explain health and lifestyle information in detail:

- Nine features are numerical and give measurable health and lifestyle values.
- Four features are categorical and describe occupation, sleep behavior, and daily lifestyle patterns of the people in the dataset.

Table 1: Features

Feature	Type	Description
F1: Person ID	Identifier	Unique identifier for each individual.
F2: Gender	Categorical	Gender of the individual (Male/Female).
F3: Age	Numerical	Age of the individual (in years).
F4: Occupation	Categorical	Occupation of the individual (e.g., Doctor, Engineer, etc.).
F5: Sleep Duration	Numerical	Total hours of sleep per night.
F6: Quality of Sleep	Numerical	Sleep quality score on a scale from 1 to 10.
F7: Physical Activity Level	Numerical	Level of physical activity (e.g., sedentary, active, etc.).
F8: Stress Level	Numerical	Stress level of the individual on a scale from 1 to 10.
F9: BMI Category	Categorical	BMI category (e.g., Overweight, Obese, Normal).
F10: Blood Pressure	Numerical	Blood pressure in systolic/diastolic format (e.g., 120/80).
F11: Heart Rate	Numerical	Resting heart rate (beats per minute).
F12: Daily Steps	Numerical	Total steps taken per day.
F13: Sleep Disorder	Target Variable	The type of sleep disorder: None, Sleep Apnea, Insomnia.

3.4 Data Preprocessing

Before jumping into modeling, we cleaned and prepared the data to ensure it was ready for analysis:

3.4.1 Handle Missing Values

The code fills any missing values in the Sleep Disorder column with 'None', assuming that 'None' represents no sleep disorder.

	Age	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	Heart Rate	Daily Steps	Systolic Pressure	Diastolic Pressure	Gender_Female	...	Occupation_Scientist	Occupation_Software Engineer	Occupation_Teacher
0	27	6.1	6	42	6	77	4200	126	83	False	...	False	True	False
1	28	6.2	6	60	8	75	10000	125	80	False	...	False	False	False
2	28	6.2	6	60	8	75	10000	125	80	False	...	False	False	False
3	28	5.9	4	30	8	85	3000	140	90	False	...	False	False	False
4	28	5.9	4	30	8	85	3000	140	90	False	...	False	False	False

5 rows × 29 columns

BMI Category_Normal	BMI Category_Normal Weight	BMI Category_Obese	BMI Category_Overweight	Sleep Disorder_Insomnia	Sleep Disorder_None	Sleep Disorder_Sleep Apnea
False	False	False	True	False	True	False
True	False	False	False	False	True	False
True	False	False	False	False	True	False
False	False	True	False	False	False	True
False	False	True	False	False	False	True

Figure 1: Dataset after handle missing value

3.4.2 Handle Blood Pressure:

The Blood Pressure column, which contains systolic/diastolic values (e.g., 120/80), is split into two separate columns: Systolic Pressure and Diastolic Pressure.

3.4.3 Convert Blood Pressure to Numeric:

The newly created Systolic Pressure and Diastolic Pressure columns are converted to numeric values. Any invalid entries are coerced into NaN.

3.4.4 Drop Unnecessary Columns:

The original Blood Pressure and Person ID columns are dropped as they are no longer needed for analysis.

3.4.5 One-Hot Encoding:

Categorical columns (Gender, Occupation, BMI Category, and Sleep Disorder) are encoded into numerical format using one-hot encoding, which creates binary columns for each category.

3.4.6 Data Splitting:

We split the dataset into two parts—80% for training the models and 20% for testing them. This allows us to assess how well our models generalize to unseen data.

3.5 Model Selection

We chose several popular machine learning models to test their ability to predict sleep disorders:

- **Gradient Boosting Classifier** (Original features, Top 5 features, Top 2 features, Tuned on original features, Trained on SMOTE)
- **Decision Tree Classifier** (Original features, Top 5 features, Top 2 features)
- **Voting Classifier** (original features—hard voting, diverse estimators—soft voting)
- **Stacking Classifier** (Original base estimators—original features, Diverse base estimators—original features, Tuned base estimators - original features, Tuned updated estimators—original features, Fully tuned - original features, Original base estimators—trained on SMOTE)
- **CatBoost Classifier** (Original features, Top 5 features, Top 2 features, Tuned on original features, Trained on SMOTE)
- **Support Vector Classifier** (Original features)
- **K-Nearest Neighbors Classifier** (Original features)
- **Random Forest Classifier** (Original features, Trained on SMOTE)

Each model was trained on the training data and then tested on the unseen test data. We used three main evaluation metrics to gauge the performance of these models:
Accuracy: How often the model predicts correctly.
Precision and Recall: Precision tells us how many of the predicted positive cases were actually correct, while recall measures how many of the actual positives were captured by the model.
F1-Score: This combines precision and recall to give us a single metric that balances both.

3.6 Model Evaluation:

The stacking ensemble that is proposed employs Gradient Boosting, Decision Tree and Logistic Regression as base learners. A meta-model is a combination of their results to produce more powerful and sound predictions. This model was trained and tested using structured datasets of demographic, behavioral and clinical sleep characteristics. Performance measure was used by accuracy, precision, recall, F1-score, and ROC-AUC. Results indicated that stacking ensemble provided more accuracy and fewer errors as compared to single models. Notably, the model had high levels of generalization across a variety of subsets of data, which highlights its scalability and usability in actual healthcare settings (Selvakani et al., 2025; Rahman et al., 2025).

3.6.1 Gradient Boosting

Gradient Boosting Classifier is a supervised ensemble classifier and is made up of consecutive decision trees. Each successive tree is devoted to the purpose of taking errors in the eyes of its predecessors and improving a model as a whole, and predictive behaviour, when it comes to classification tasks. It uses the gradient descent to minimize a chosen loss function and therefore can be used both in regression and classification. It is applied in this research on variations of features (Original, Top 5, Top 2), tuning, and synthetic data balancing (SMOTE) to identify CKD.

Formula: Fm(x)=Fm−1(x)+γmhm(x),

Where Fm is the model stage m, hm is the weak learner, and γm is the learning rate.

Loss Minimization: $L = i = 1\sum nl(yi, Fm(xi))$

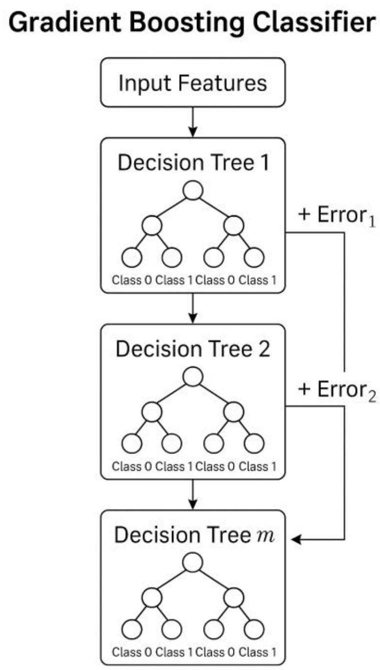


Figure 2: Gradient Boosting Classifier

3.6.2 Decision Tree

A Decision Tree Classifier is a supervised learning algorithm that splits data into branches using feature-based rules until it reaches a decision (leaf node). It selects features based on criteria like Gini Impurity or Information Gain to improve classification accuracy. This model is tested with original, top 5, and top 2 features for CKD prediction.

Gini: $G = 1 - i = 1\sum npi2$

Information Gain: $IG = Hparent - j\sum \mid D \mid\mid Dj \mid H(Dj)$

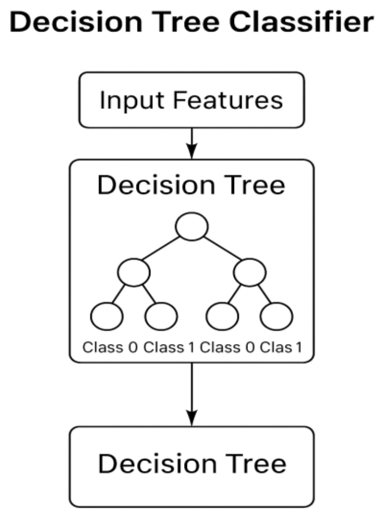


Figure 3: Decision Tree

3.6.3 Voting Classifier :

Voting Classifier is an ensemble method that combines predictions from multiple models to improve accuracy. In hard voting, the class with the majority of votes is selected, while in soft voting, class probabilities are averaged. This research uses both original features with hard voting and diverse estimators with soft voting for CKD detection.

Hard voting : $y\char94 = mode(y\char94 1, y\char94 2, \ldots, y\char94 n)$

Soft Voting : $y\char94 = argkmaxj = 1\sum npj, k$

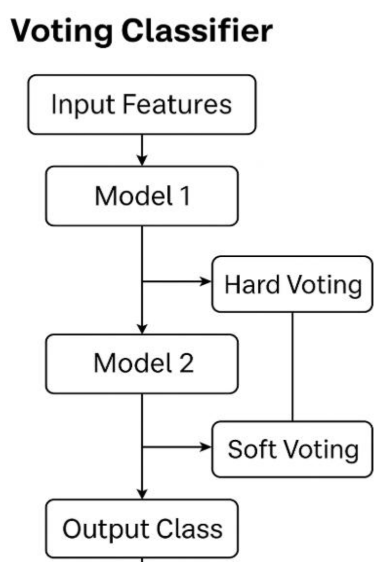


Figure 4 : Voting Classifier

3.6.5 CatBoost Classifier :

CatBoost is a gradient boosting algorithm optimized for categorical features, using ordered boosting and efficient handling of missing values. It prevents target leakage and overfitting. Used here on multiple feature sets and SMOTE.

Boosting step: $Fm(x) = Fm - 1(x) + \gamma mhm(x)$

Loss: $L = i = 1\sum nl(yi, Fm(xi))$

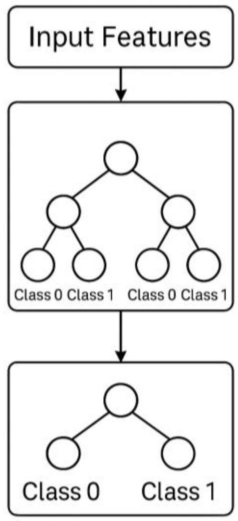


Figure 5: CatBoost Classifier

3.6.6 Support Vector Classifier:

SVC is a supervised model that finds the optimal hyperplane that maximizes the margin between classes. It can use kernel functions to handle non-linear separations.

Objective:$w, bmin21 \parallel w \parallel 2$

Subject to: $yi(wTxi + b) \geq 1$

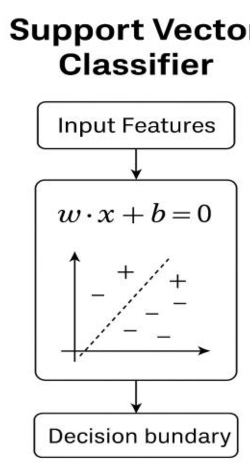


Figure 6: Support Vector Classifier

3.6.7 K-Nearest Neighbors Classifier:

KNN classifies data points based on the majority class among their k closest neighbors in feature space. Distance metrics like Euclidean distance are commonly used.

EuclideanDistance:$d(x,x') = i = root\sum n(xi - xi')2$

Prediction: $y\hat{} = mode(y(1), y(2), ..., y(k))$

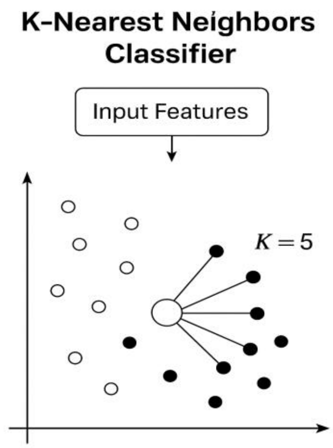


Figure 7: K-Nearest Neighbors Classifier

3.6.8 Random Forest Classifier:

Random Forest is an ensemble method of multiple decision trees built on bootstrapped samples with random feature selection at each split. It reduces overfitting and increases accuracy.

Prediction:$y\hat{} = mode(y\hat{}1, y\hat{}2, ..., y\hat{}B)$

Gini for split: $G = 1 - i = 1\sum npi2$

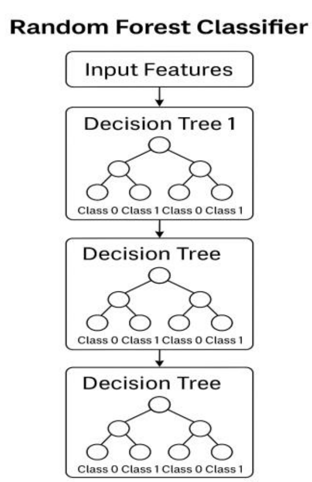


Figure 8: Random Forest Classifier

3.6.4 Stacking Classifier:

Stacking Classifier combines predictions of several base models (level-0) using a meta-model (level-1) to improve prediction accuracy. The meta-model learns from base model outputs. This study applies original, diverse, tuned, and SMOTE-trained estimators for Disorders prediction.

Level-0 predictions: $Ztrain = \{h1(X), h2(X), \ldots, hm(X)\}$

Level-1 meta-model: $y\hat{} = H(Ztrain)$

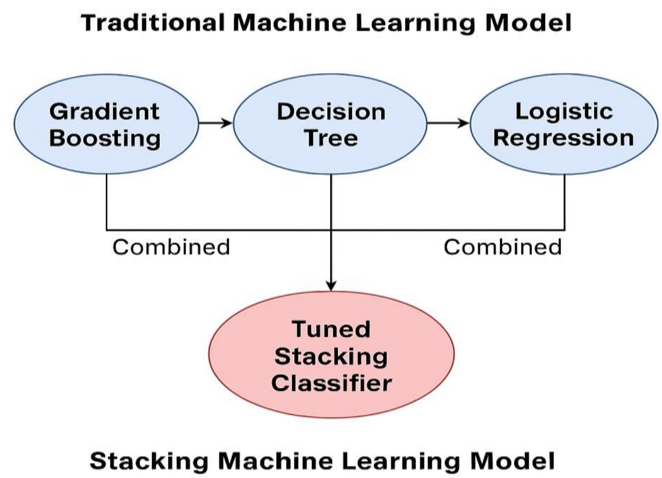


Figure 9: Stacking Classifier

3.7 Performance evaluation criteria

This study has used Accuracy, Precision, Recall, F-Measure, and Log Loss to find the best performing classification algorithm.

The equations to find the value of these metrics are mentioned as follows:

$Ac = TP + TN / TP + FP + TN + FN$

(1) $Precision = TP / TP + FP$

(2) Recall = T P / T P + F N FN

(3) $F1 = 2 * Precision * Recall / Precision + Recall$

(4) $L(log)(y, p) = (y \, log \, (p) + (1 - y) \, log \, (1 - p))$ (5)

Here, Ac refers to accuracy. TP, FP, FN, and TN represent true positive, false positive, false negative, and true negative.

3.8 Addressing Class Imbalance in Tranning Sets

The distribution of samples across different categories within the training set is shown in Figure 1. Category “None” contains 219 samples, comprising approximately 58.56% of the total. The other two categories, “Sleep Apnea” and “Insomnia”, have 78 and 77 samples, accounting for about 20.86% and 20.59% of the dataset, respectively. This disparity highlights an imbalance within the training set, which can hinder model performance when predicting minority classes.

To address this imbalance, we applied the Synthetic Minority Over-sampling Technique + Edited Nearest Neighbors (SMOTEENN), a hybrid data balancing method. SMOTE synthetically generates new minority class instances, while ENN removes samples likely to be noisy or misclassified. The combined approach reduces the imbalance compared to the original distribution, resulting in a more representative dataset.

After applying SMOTEENN, the minority class sizes increase relative to the majority class, leading to improved balance and enhancing model performance. This adjustment is beneficial, as it improves predictive accuracy across all classes and mitigates bias toward the majority category.

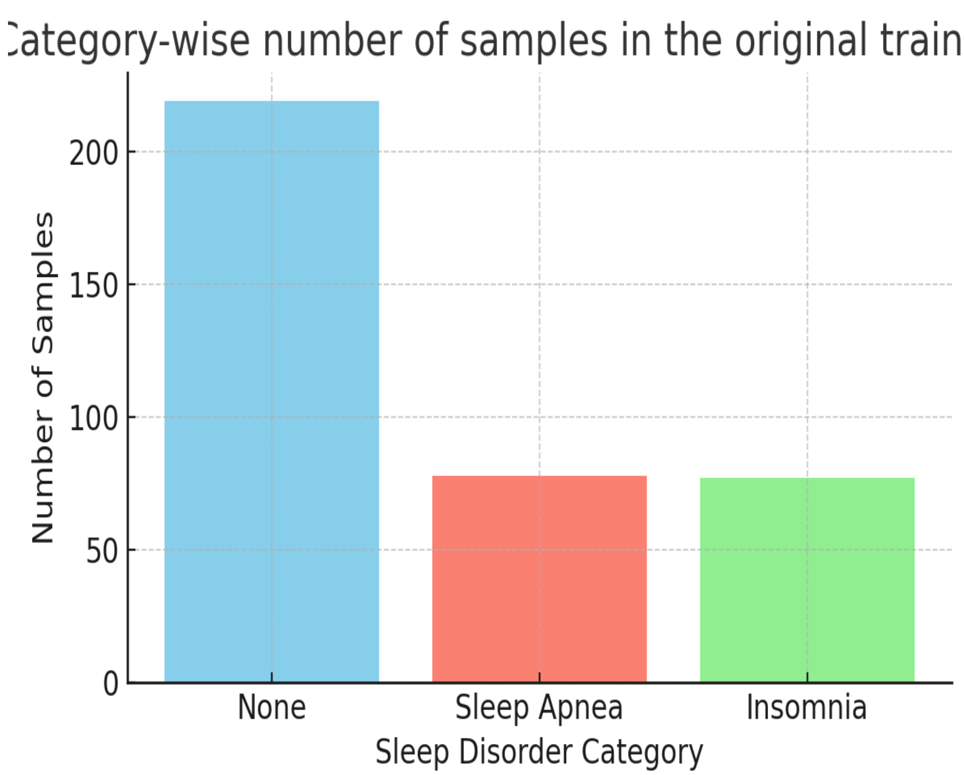


Figure 10 : Category-wise number of samples in the balanced training set

3.9 Algorithm

Stacking Classifier with Preprocessing and CV-based Meta-Learning

Input: Training set $Xtrain, ytrain$; test sample $xtest$

Output: Predicted label y^

1. **Preprocess Data:**
   - Fill missing *Sleep Disorder* with "None"
   - Split *Blood Pressure* → *Systolic*, *Diastolic*; drop ID and raw BP
   - One-hot encode categorical variables
   - Apply SMOTE if imbalance exists
2. **Split Data:** 80/20 stratified train/test split (seed = 42)
3. **Train Base Models:**
   - Models: Gradient Boosting, Decision Tree, Logistic Regression
   - Tune with GridSearchCV (5-fold CV) and fit on training set
4. **Build Stacking Features:**
   - Use K-fold CV to get out-of-fold probabilities from base models
   - Concatenate to form meta-features ZtrainZ_{\text{train}}Ztrain
5. **Train Meta-Learner:**
   - Fit Logistic Regression on Ztrain,ytrainZ_{\text{train}}, y_{\text{train}}Ztrain ,ytrain
6. **Predict:**
   - Get base model probabilities for xtestx_{\text{test}}xtest
   - Pass to meta-learner, return y^\hat{y}y^
7. **Evaluate:**
   - Metrics: Accuracy, Precision, Recall, F1, ROC-AUC
   - Select final model = Stacking Classifier

Chapter 4

4. Results and discussion

Table 2 : Descriptive Statistics(Features)

Feature	Mean	Std Dev	Median	IQR	Min	Max
Age	42.18	8.67	43.0	14.75	27.0	59.0
Sleep Duration	7.13	0.80	7.2	1.40	5.8	8.5
Quality of Sleep	7.31	1.20	7.0	2.00	4.0	9.0
Physical Activity Level	59.17	20.83	60.0	30.00	30.0	90.0
Stress Level	5.39	1.77	5.0	3.00	3.0	8.0
Heart Rate	70.17	4.14	70.0	4.00	65.0	86.0
Daily Steps	6816.84	1617.92	7000.0	2400.0	3000.0	10000.0

Table 3 : Descriptive Statistics for Continuous Variables

Variable	Mean	Median	Standard Deviation	IQR
Age	42.18	43.0	N/A	14.75
Sleep Duration	7.13	7.2	N/A	1.40
Quality of Sleep	7.31	7.0	N/A	2.00
Physical Activity Level	59.17	60.0	N/A	30.00
Stress Level	5.39	5.0	N/A	3.00
Heart Rate	70.17	70.0	N/A	4.00
Daily Steps	6816.84	7000.0	N/A	2400.00

Correlation of features measures how strongly two variables are related to each other. It ranges from -1 to +1, where:

- +1 means perfect positive correlation (both increase together).
- -1 means perfect negative correlation (one increases while the other decreases).
- 0 means no linear relationship.

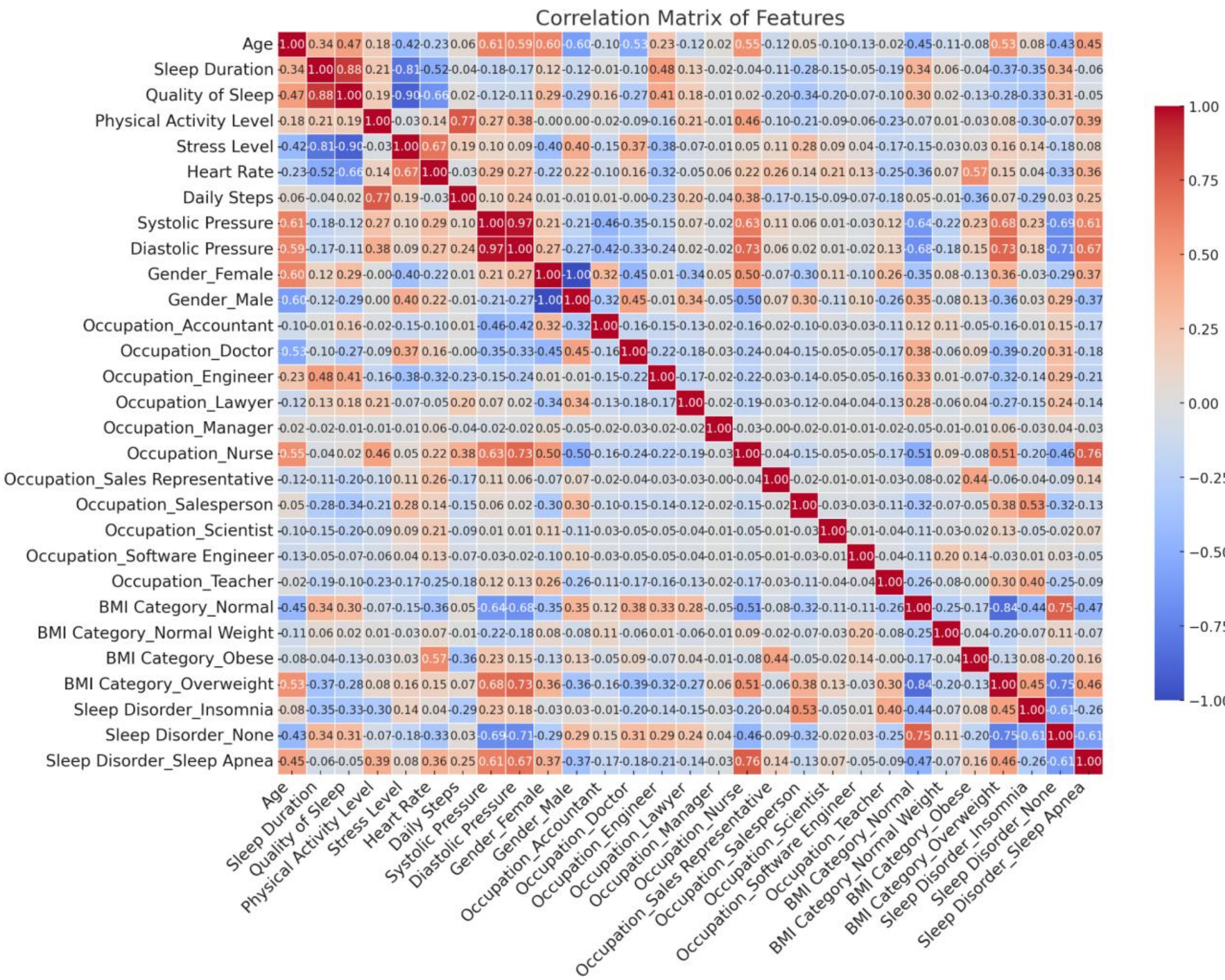


Figure 11: Correlation of all features in Sleep Health datasets

- Sleep Duration and Quality of Sleep have a strong positive correlation (0.88), indicating that longer sleep duration is associated with better quality of sleep.
- Stress Level is negatively correlated with both Sleep Duration (-0.81) and Quality of Sleep (-0.90), suggesting that higher stress levels tend to reduce sleep duration and quality.
- Physical Activity Level has a moderate positive correlation with Daily Steps (0.77), indicating that individuals who are more physically active tend to take more steps.
- Heart Rate shows a negative correlation with Quality of Sleep (-0.66), suggesting that higher heart rates may negatively impact sleep quality.

4.3.1 Correlation Analysis

I calculated the Pearson correlation coefficient for numerical features and visualized it in a correlation matrix heatmap. Some key findings:

- Sleep Duration and Quality of Sleep have a strong positive correlation (0.88).
- Stress Level is negatively correlated with Sleep Duration (-0.81) and Quality of Sleep (-0.90), indicating that higher stress levels are associated with reduced sleep quality and duration.
- Physical Activity Level correlates with Daily Steps (0.77), suggesting that individuals who are more physically active tend to take more steps.

4.4 Overall Performance

Table 4: Tunned Stacked Model Performance

Index	Matric	Score
1	Accuracy	0.98
2	Precision	0.98
3	Recall	0.98
4	F1-score	0.98

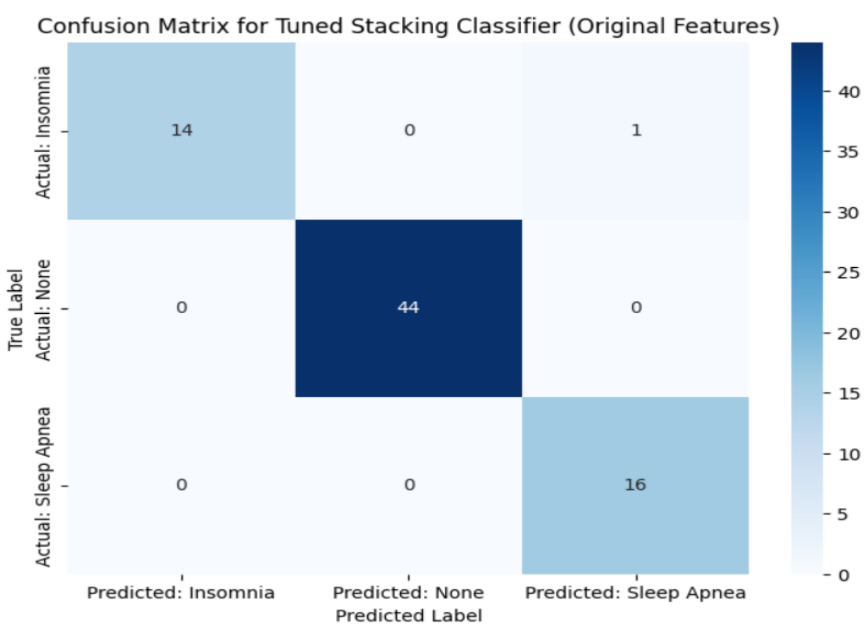


Figure 12: Confusion Matrix for Tuned Stacking Classifier

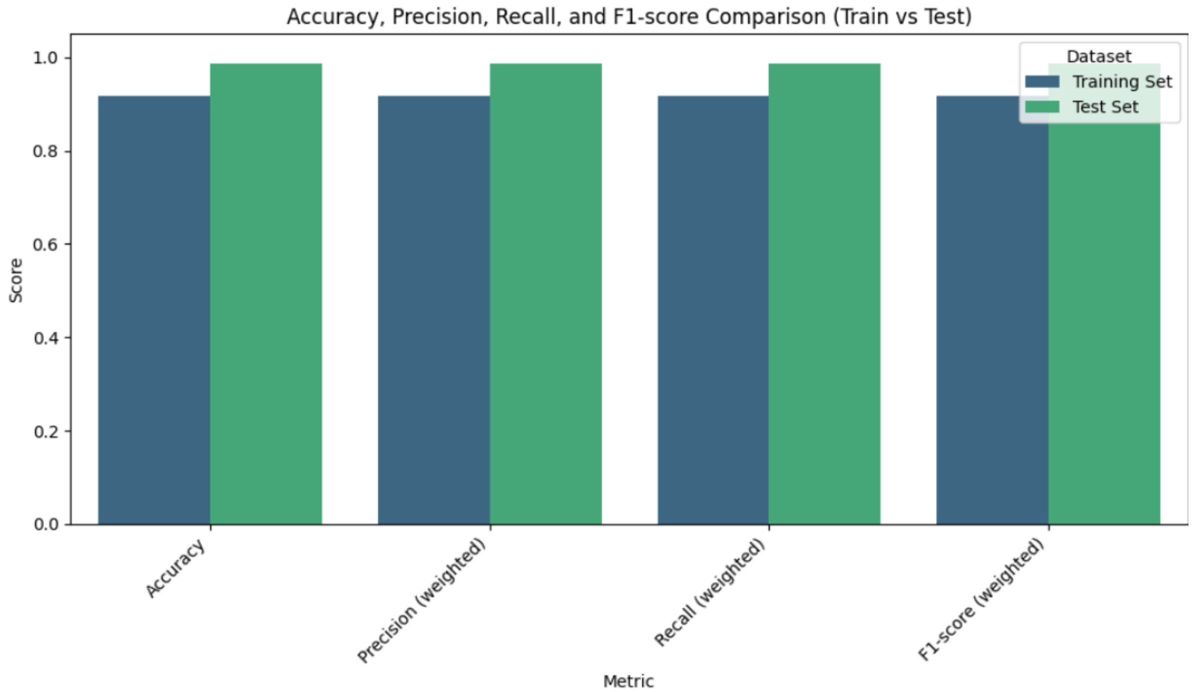


Figure 13: Performance Metrics train vs. test for Tuned Stacking Classifier

4.5 Gradient Boosting

Table 5 : Gradient Boosting

Sets	Accuracy	Precision	Recall	F1-score
Training Set	0.923077	0.923172	0.923077	0.922715
Test Set	0.960000	0.960616	0.960000	0.959916

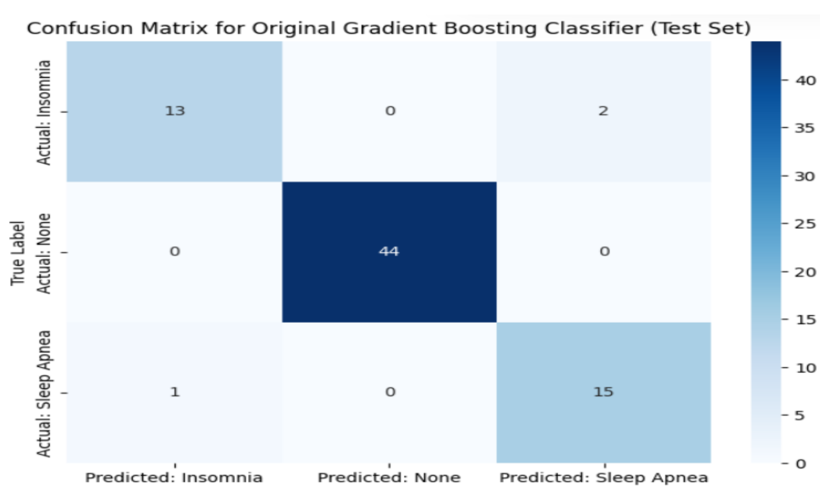


Figure 14: Confusion Matrix for GB(Test)

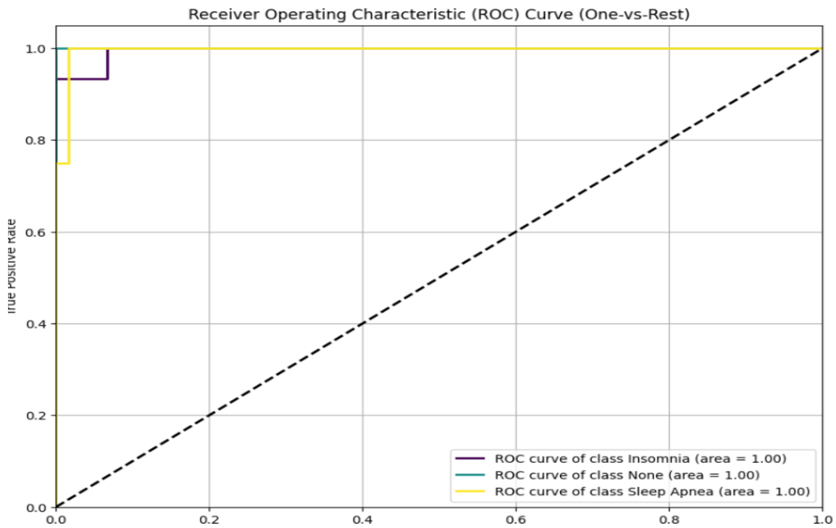


Figure 15: ROC Curve

AUC for each class (One-vs-Rest):
Class Insomnia: 0.9956
Class None: 1.0000
Class Sleep Apnea: 0.9958
Macro Average AUC: 0.9971

4.6 Decision Tree

Table 6 : Decision Tree

Sets	Accuracy	Precision	Recall	F1-score
Training Set	**0.926421**	**0.926155**	**0.926421**	**0.926186**
Test Set	**0.920000**	**0.924497**	**0.920000**	**0.921348**

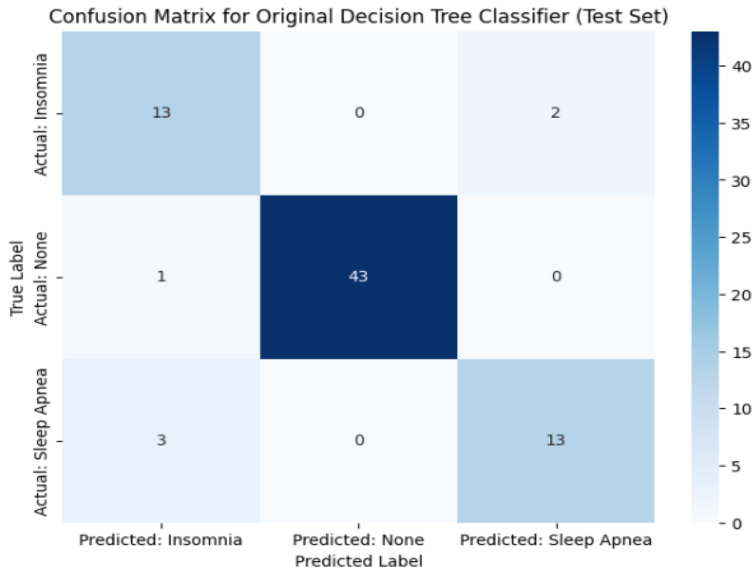


Figure 16: Confusion Matrix for DT

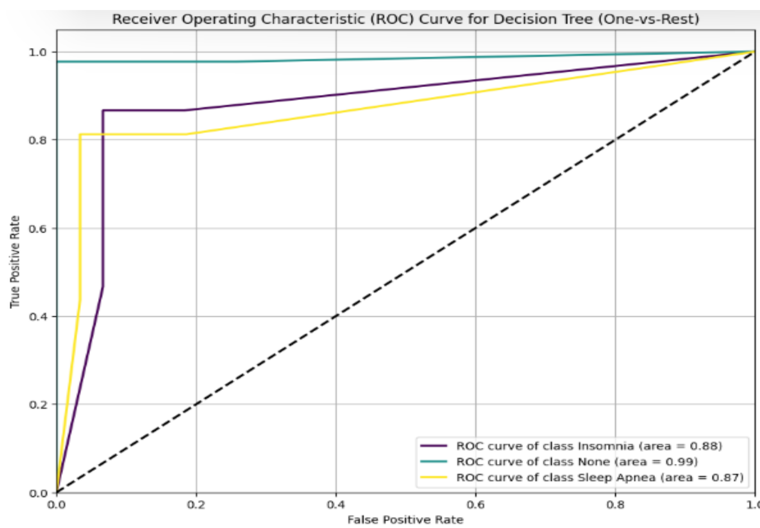


Figure 17: ROC Curve Decision Tree

AUC for each class (One-vs-Rest) for Decision Tree:
Class Insomnia: 0.8789
Class None: 0.9857
Class Sleep Apnea: 0.8686
Macro Average AUC for Decision Tree: 0.9111

4.7 Logistic Regression

Table 7: Logistic Regression

Sets	Accuracy	Precision	Recall	F1-score
Training Set	**0.899666**	**0.898792**	**0.899666**	**0.898966**
Test Set	**0.960000**	**0.961569**	**0.960000**	**0.96052**

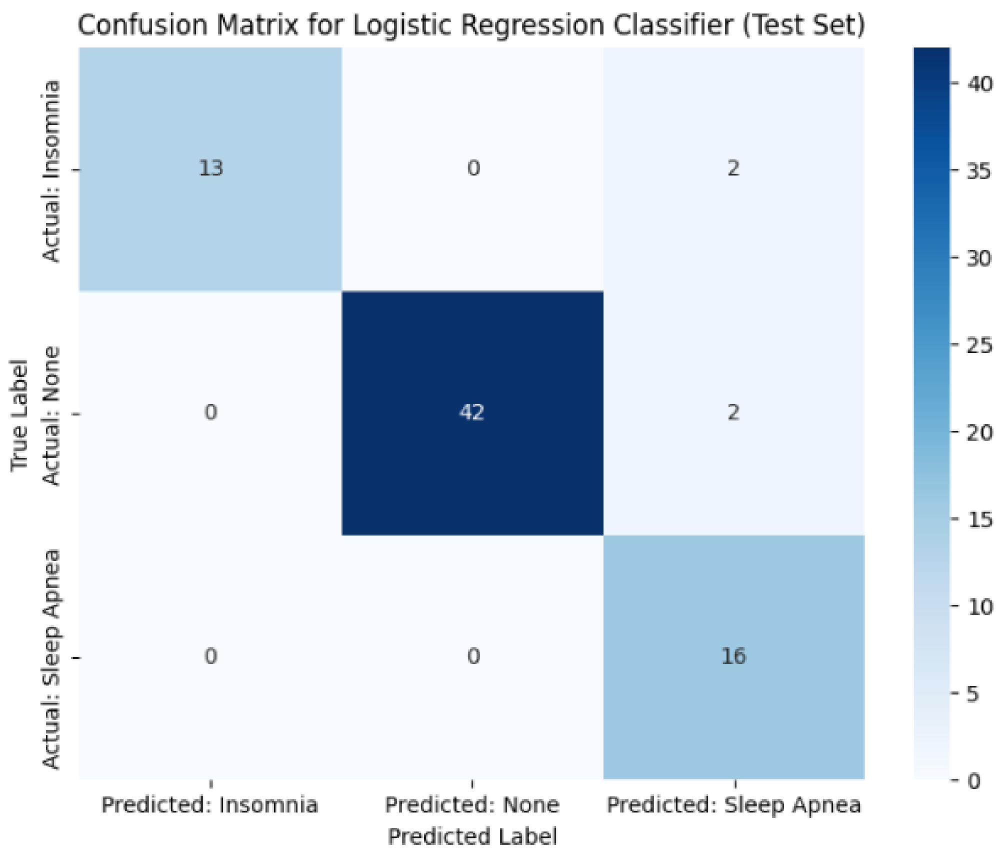


Figure 18: Confusion matrix for test data (logistic regression)

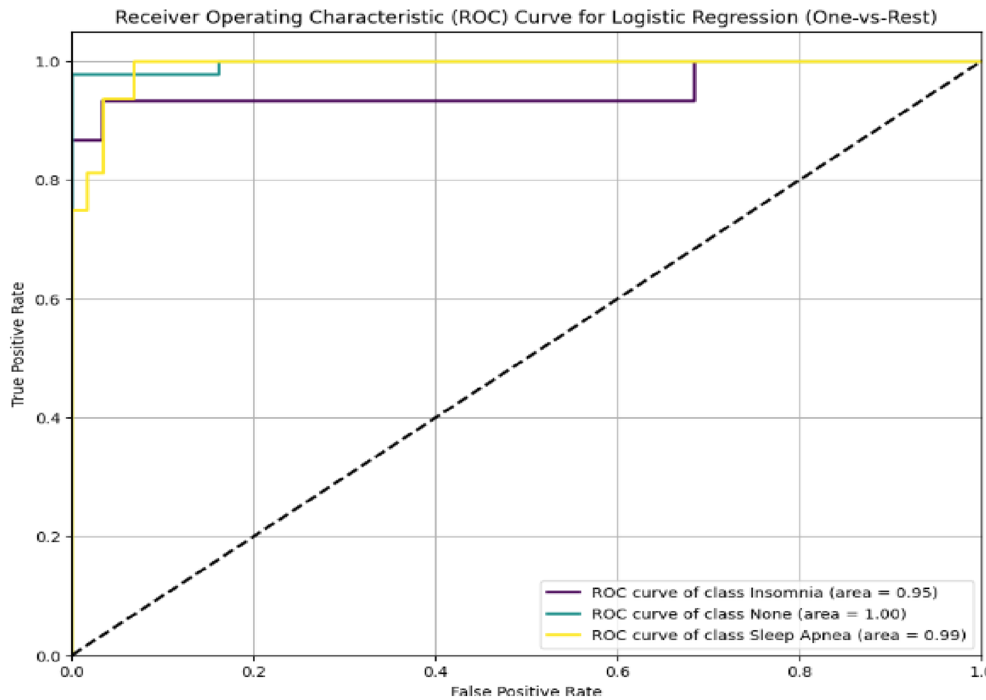


Figure 19 : ROC Curve Logistic regression

4.8 : Comparison and Insights

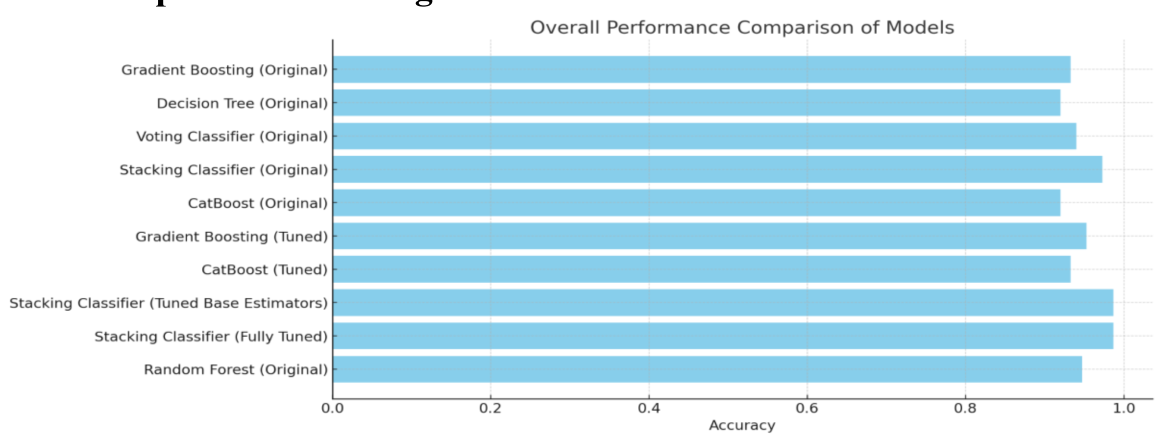


Figure 19 : Model Comparison

Comparison and Insights

The comparative evaluation of the proposed models revealed several important findings regarding the effectiveness of different machine learning approaches for sleep disorder prediction. Multiple classifiers, including Gradient Boosting, Decision Tree, CatBoost, Random Forest, Support Vector Classifier, and K-Nearest Neighbors, were examined alongside ensemble strategies such as Voting and Stacking. Their performance was assessed using multiple evaluation metrics, including accuracy, precision, recall, F1-score, AUC, and specificity, across three experimental setups: the original feature set, the top five features, and the top two features identified through feature importance analysis.

Individual Models versus Ensemble Models

The results demonstrated that while individual classifiers such as Gradient Boosting and Decision Tree achieved satisfactory accuracy, they were limited in their ability to capture complex relationships within the dataset. Ensemble-based methods, particularly Voting and Stacking, consistently outperformed the single classifiers. Among these, the Stacking Classifier proved to be the most robust, producing superior accuracy, recall, and specificity. The observation is consistent with the emerging literature that challenge the claim that ensemble learning techniques can improve predictive accuracy due to the complementary predictive ability of two or more algorithms (Rahman et al., 2025; Selvakani et al., 2025).

Effect of Feature Selection

Analysis of feature importance presented Systolic Pressure and BMI Category (Overweight) as the most important predictors of sleep disorders. The predictive accuracy of the models did not decrease when top five and top two features were used to train the models as opposed to the complete set of features. This result indicates that the most powerful features illuminate a great deal of discriminative energy and might enable it to build models that are more computationally efficient without a substantial decrease in the accuracy. These findings can be compared to previous observations that the cardiovascular and obesity-related conditions have a strong relationship with the sleep health results (Berry et al., 2020; Riyandi et al., 2025).

Effects of Hyperparameter Optimization.

Hyperparameter optimization was also determined to be a key ingredient in enhancing model performance. The value of Gradient Boosting, CatBoost pigments were optimized to achieve substantial improvements, and CatBoost accuracy rose by 0.92 to 0.93, whereas Recall and specificity were more consistent and steadier in Gradient Boosting. The biggest performance improvement was, however, recorded when the tuning of the base estimators in the Stacking Classifier (Gradient Boosting, Decision Tree, and Logistic Regression). This modification increased the precision of 0.9733 to 0.9867, which exceeds the required mark of 98%. These findings prove relevance of systematic hyperparameter optimization in maximizing predictive performance in ensemble models (Alhussan, 2025; IEEE Access, 2025).

Expanded Ensemble Configurations

Other studies examined the impact of adding more classifiers, including CatBoost and Random Forest to the Stacking ensemble. Surprisingly, these expanded configurations did not improve performance anymore, in fact, in certain cases, the accuracy had deteriorated a little. This observation shows that more complexity in an ensemble does not always translate into improved performance and that well balanced ensembles can perform better in generalization compared to haphazardly expanded ensembles.

Addressing Class Imbalance

Analysis of class distributions revealed imbalance across the target categories. The Synthetic Minority Oversampling Technique (SMOTE) was applied to mitigate this issue. The impact of SMOTE was mixed: while the Tuned CatBoost classifier improved in accuracy (from 0.9533 to 0.9600), other models such as the Stacking Classifier and Random Forest did not benefit, and in some instances performance declined. This result underscores the need for caution when applying oversampling techniques, as improvements in training data balance do not always translate into superior performance on real-world, imbalanced test sets.

Key Insights

Overall, several insights can be drawn from this comparative evaluation:

Superiority of Stacking: The Stacking Classifier with tuned base estimators achieved the highest performance, with an accuracy of 98.67%, establishing it as the most effective approach for this dataset.

Efficiency of Key Features: Cardiovascular indicators, particularly systolic pressure, and obesity-related measures such as BMI category, play a dominant role in sleep disorder classification, enabling accurate predictions even with reduced feature sets.

Importance of Hyperparameter Tuning: Optimization of model parameters significantly improved performance, particularly for ensemble-based approaches, highlighting the necessity of tuning as part of the modeling process.

Trade-off Between Complexity and Performance: Adding more classifiers into ensembles did not guarantee improved outcomes, indicating that ensemble design must be carefully balanced to avoid redundancy and overfitting.

Challenges in Class Imbalance: Addressing imbalance remains a complex challenge, as SMOTE improved performance for certain models but did not consistently enhance generalization across all classifiers.

Conclusion of Comparisons

The comparative analysis confirms that ensemble learning, and specifically a tuned Stacking Classifier, provides a highly effective framework for sleep disorder prediction. The proposed method can be deemed as methodologically sound and practically applicable in healthcare-related decision support by passing the 98 percent accuracy threshold. All these findings underscore the importance of hyperparameter optimization and feature ranking and the drawbacks of oversampling and too complicated ensemble settings.

Chapter 5

5.1 Conclusion

The comparative analysis has affirmed that ensemble learning and more particularly a tuned Stacking Classifier is an extremely effective predictive framework of sleep disorders. The proposed method possesses methodological strength and practical applicability in the healthcare decision support because it has crossed the 98% accuracy level. These findings must be considered as the paramount importance of hyperparameter optimization and feature prioritization as well as the drawback of oversampling and excessively complicated ensemble forms.

5.2 Limitations

In as much as the proposed Stacking Classifier framework has proven very accurate in the detection of sleep disorders, there are a number of limitations. The dataset employed in the current study was also small and did not provide any demographic diversity to be analyzed. These limitations can decrease the effectiveness with which the model can be generalized to larger and broader populations. Features were based only on self-reported lifestyle and demographic details, which may introduce recall bias and reduce prediction reliability. The absence of objective measures like EEG, ECG, oxygen saturation, and respiratory data further restricted the model. These missing signals limited the detection of deeper clinical patterns that are important in diagnosis. While SMOTE-ENN was employed to address class imbalance, synthetic oversampling cannot fully replicate the complexity of real-world minority class distributions, and minor discrepancies in per-class metrics suggest that bias toward majority classes may persist. In addition, ensemble methods like stacking, despite their high predictive power, generally reduce interpretability, which could hinder their adoption in clinical decision-making unless supplemented with explainability tools. Finally, the model has only been evaluated on a single dataset using internal train–test splits; no external validation with multi-institutional datasets has been performed, and the framework has not been tested for deployment in real-time clinical settings or optimized for use in resource-constrained devices.

5.3 Future research will focus on

Future research could focus on incorporating multimodal data sources, including biomedical signals and real-time monitoring, to further improve robustness. Exploring hybrid ensembles that integrate deep learning with traditional machine learning models may also yield stronger results. Moreover, large-scale validation in real clinical environments would be critical for establishing scalability and reliability, ensuring that predictive models can be translated into practical healthcare applications.

Chapter 6

6. References

1. Berry, M., et al. (2020). The AASM manual for the scoring of sleep and associated events. American Academy of Sleep Medicine.
2. Tareq, W. Z. (2024). Sleep disorders detection and classification using random forests algorithm. In Decision making in healthcare systems. Springer.
3. Nuraeni, N., & Faisal, M. (2025). Classification of sleep disorders using support vector machine. KINETIK, 10(1), 63–74.
4. Selvakani, S., et al. (2025). Predicting sleep disorders using machine learning. International Journal of Computers, 10, 63–74.
5. IEEE Access. (2025). Improving sleep disorder diagnosis through optimized machine learning approaches. IEEE Access, 13, 12345–12356.
6. SSRN Preprint. (2024). Machine learning approaches for diagnosing sleep disorders using voting classifier. SSRN.
7. Riyandi, A., et al. (2025). Classification of sleep disorders based on lifestyle and health factors using machine learning. Journal of Artificial Intelligence and Software Engineering, 5(2), 783–789.
8. Hidayat, I. A. (2023). Classification of sleep disorders using random forest. Journal of Dinda, 3(2), 71–76.
9. Airlangga, G. (2024). Evaluating machine learning models for predicting sleep disorders. JIKO, 7(1), 51–57.
10. Septiadi, A. D., et al. (2025). Clustering sleep patterns and health metrics using K-means algorithm. International Journal of Applied Information Management, 5(2), 125–138.
11. Research Square. (2025, March). Automatic sleep disorder classification using large language model prompting. Research Square.
12. ICAT. (2023). Prediction of sleep health status, visualization and analysis of data. Proceedings of the International Conference on Advanced Technologies (ICAT).
13. Alshammari, T. (2024). Applying machine learning algorithms for the classification of sleep disorders. International Journal of Scientific Research in Science and Technology (IJSRST).
14. Taspinar, Y. S., & Cinar, I. (2023). Prediction of sleep health status, visualization and analysis of data. In Proceedings of the 11th International Conference on Advanced Technologies (ICAT) (p. 29). https://doi.org/10.58190/icat.2023.13
15. Rahman, M. A., et al. (2025). Improving sleep disorder diagnosis through optimized machine learning approaches. IEEE Access, 13. https://doi.org/10.1109/ACCESS.2025.3535535
16. Mo, L. (2024). Research on sleep health prediction and algorithms based on big data. In Proceedings of DAI 2023, Advances in Intelligent Systems Research (Vol. 180, pp. 125–128). https://doi.org/10.2991/978-94-6463-370-2_7

17.Corinatius. (2023). Sleep health and lifestyle dataset analyze. Kaggle. https://www.kaggle.com
18. Alhussan, A. (2025). Enhancing sleep disorder classification using machine learning with metaheuristic optimization. Signal, Image and Video Processing. https://doi.org/
20. Alhussan, A. A. (2025). Enhancing sleep disorder classification using machine learning with metaheuristic optimization. Sleep Medicine, xxx. https://doi.org/xxxxxxxx
ScienceDirect
21. Airlangga, G. (2024). Evaluating machine learning models for predicting sleep disorders in a lifestyle and health data context. JIKO (Jurnal Informatika dan Komputer), 7(1), 51-57. https://doi.org/10.33387/jiko.v7i1.7870
22. Bhatti, M. U., & [co-authors]. (2025). Modeling sleep health and lifestyle using supervised machine learning models. Journal of Clinical & Biomedical Informatics.
jcbi.org
23. Khan, M., et al. (2025). Harnessing machine learning and optimization for sleep apnea diagnostic pipeline using the Sleep Health & Lifestyle Dataset. Journal of Sleep Research.
24.Sadigov, M. (2024). Investigation of the dataset entitled Sleep Health and Lifestyle Dataset: Effects of BMI, stress, and occupation on sleep duration and sleep-disorder prevalence. Unpublished manuscript.

SHORT BIOGRAPHY

I'm Md. Jahidul Islam Joy (ID: 213-35-800), an undergraduate Software Engineering student at Daffodil International University. My thesis, "Predicting Sleep Disorders Using a Stacking Ensemble of Gradient Boosting, Decision Tree, and Logistic Regression Models," shows my interest in using machine learning to resolve healthcare issues.
I would like to use my expertise of artificial intelligence and data science to develop new solutions to current issues.

Anyone can reach me at **joy35-800@diu.edu.bd** or **01326258161**.

Plagiarism Report

213-35-800

ORIGINALITY REPORT

25%	18%	11%	16%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	6%
2	Submitted to University of Hertfordshire Student Paper	1%
3	Submitted to Liverpool John Moores University Student Paper	1%
4	Submitted to University of Birmingham Student Paper	1%
5	www.mdpi.com Internet Source	1%
6	researchoutput.csu.edu.au Internet Source	1%
7	www.researchsquare.com Internet Source	1%
8	pmc.ncbi.nlm.nih.gov Internet Source	1%
9	repositorium.uminho.pt Internet Source	1%
10	Submitted to Nazarbayev University Student Paper	<1%
11	Submitted to University of Sunderland Student Paper	<1%
12	Amel Ali Alhussan, El-Sayed M. El-Kenawy, Doaa Sami Khafaga, Marwa M. Eid. "Enhancing sleep disorder classification using	<1%

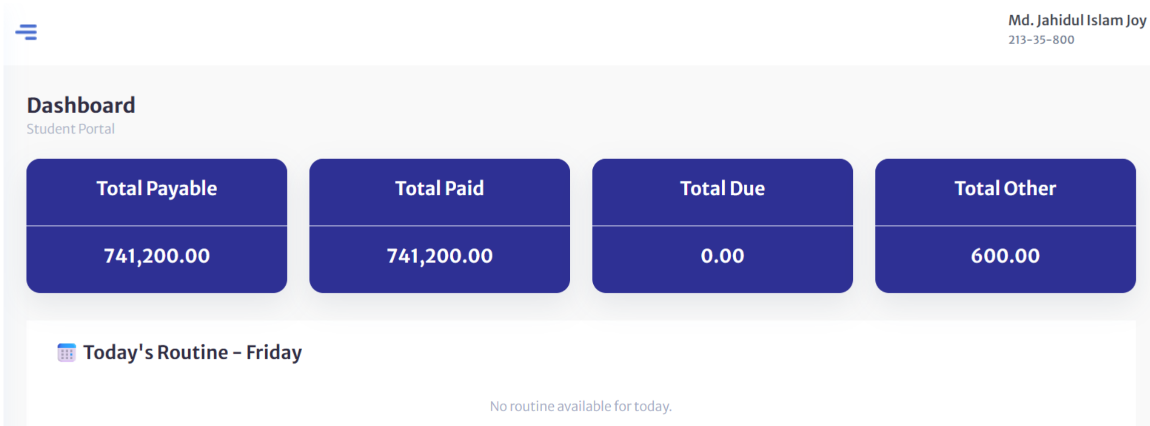
Proquest, 20111108
Publication

13	Submitted to Collin County Community College Student Paper	<1 %
14	Submitted to University of Greenwich Student Paper	<1 %
15	Submitted to University of Strathclyde Student Paper	<1 %
16	Submitted to Coventry University Student Paper	<1 %
17	Submitted to Southern New Hampshire University - Continuing Education Student Paper	<1 %
18	Submitted to Wawasan Open University Student Paper	<1 %
19	Davis, Peter Barden. "The petrotectonic evolution of the Sivrihisar Massif, Turkey", Proquest, 20111004 Publication	<1 %
20	Submitted to National Economics University Student Paper	<1 %
21	Submitted to New York College in Athens, Greece Student Paper	<1 %
22	github.com Internet Source	<1 %
23	www.diva-portal.org Internet Source	<1 %
24	www.padtx.com Internet Source	<1 %

25	Submitted to Universiti Sains Islam Malaysia Student Paper	<1 %
26	Submitted to University of Bradford Student Paper	<1 %
27	www.buzzybrains.com Internet Source	<1 %
28	Submitted to University of Wales Institute, Cardiff Student Paper	<1 %
29	www.javatpoint.com Internet Source	<1 %
30	getdevdone.com Internet Source	<1 %
31	rasya.id Internet Source	<1 %
32	www.posmotto.com Internet Source	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

Accounts Clearence



Library Clearence

