Sleep Stage Classification Using EEG Data and Machine Learning

A Master's Report

Submitted in partial fulfillment of the requirement for the award of the degree of

Master of Technology in Artificial Intelligence

by

Tanmay Rathod

23MAI007

Under the guidance of

Dr. SANTOSH SATAPATHY

Department of Information and Communication Technology

School of Technology

Pandit Deendayal Energy University

Gandhinagar – 382426. Gujarat - India

August, 2024

Approval Sheet

This thesis entitled **"Sleep Stage Classification Using EEG Data and Machine Learning"** by **Tanmay Rathod** is recommended for the degree of **M.Tech** in **Artificial Intelligence**.

Dr SANTOSH SATAPATHY

Supervisor

Dr. Paawan Sharma,

HoD, ICT

Date: ______________________

Place: ______________________

Student Declaration

I, **Tanmay Rathod**, hereby declare that this written submission represents my ideas in my own words, and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated, or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Pandit Deendayal Energy University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

____

Tanmay Rathod

Roll No: 22MAI007

Date: ________________________

Acknowledgment

While my name appears as the sole contributor to the completion of this summer internship, it is important to recognize that the guidance and support of many individuals played a significant role in its success. This internship is the result of a collective effort, shaped by the contributions of numerous people, to whom I am deeply grateful.

I extend my heartfelt appreciation to my internal guide, Dr. SANTOSH SATAPATHY, for his consistent guidance, encouragement, and insightful suggestions throughout this internship. His belief in my abilities provided the confidence and motivation necessary to complete this project. Dr. SANTOSH SATAPATHY's support, along with his generous provision of time and access to resources, was instrumental in achieving the objectives of this internship.

I am also profoundly thankful for the unwavering support of my parents. Their continuous inspiration, moral backing, and blessings have been the cornerstone of my journey. Their understanding and encouragement have been invaluable, and I am truly fortunate to have had their steadfast support throughout this endeavor.

Tanmay Rathod

Abstract

This research focuses on the classification of sleep stages using EEG data and machine learning algorithms. The objective is to develop a model that can accurately classify different sleep stages, aiding in the diagnosis of sleep disorders. The study compares various machine learning techniques and discusses the results obtained from the experiments.

This project aims to design a machine learning model that could accurately classify sleep stages using EEG data. The research work focuses on EEG signal preprocessing, approaches to handling class imbalance through resampling, and applying various machine learning algorithms to achieve high accuracy in sleep stage classification. Sleep is categorized into multiple stages—N1, N2, N3, N4, and REM—each distinguished by identifying different patterns in EEG signals.

A key consideration in this research is class imbalance, where some sleep stages are underrepresented, potentially affecting model performance. EEG data was segmented into 30-second epochs, labeled according to the relevant sleep stage, and then preprocessed into feature vectors for model input. SMOTE (Synthetic Minority Over-sampling Technique) was employed to generate synthetic samples for underrepresented classes, thereby balancing the dataset.

The models were trained using Machine Learning Techniques, and their performance was evaluated based on accuracy, ROC-AUC metrics, and log loss. An iterative training approach ensured continuous improvement, resulting in a robust model capable of reliably classifying sleep stages.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Sleep is an important part of life. It provides us with the ability to learn, maintain attention to day-to-day activities, and support mental health. During sleep, the body's major organs can coordinate with one another, which impacts our overall health. A lack of sleep can lead to serious health issues such as insomnia. Sleep also plays a role in replenishing muscle glycogen, which helps in muscle recovery. Research indicates that we spend approximately one-third of our lives sleeping.

1.2 Brain Waves and Sleep Stages

1.2.1 Brain Wave Characteristics

Our brain produces bio-signals as a result of electrical activity. These electrical voltages range to a few millionths of a volt.

TABLE 1.1
Brain States and Corresponding Frequency Bands

Frequency Band	Frequency	Brain States
Gamma (γ)	> 35 Hz	Concentration
Beta (β)	$12 - 35$ Hz	Anxiety dominant, active, external attention, relaxed
Alpha (α)	$8 - 12$ Hz	Very relaxed, passive attention
Theta (θ)	$4 - 8$ Hz	Deeply relaxed, inward-focused
Delta (δ)	$0.5 - 4$ Hz	Sleep

1.2.2 Sleep Data Extraction Techniques

The American Academy of Sleep Medicine (AASM) has investigated data extraction techniques in several domains:

1. Time
2. Frequency
3. Time and Frequency
4. Nonlinear and Entropy Domain

With advancements in signal processing, statistical analysis, and computer science, there is a gap between machine learning and sleep staging automation. Barriers to adoption by sleep data professionals include the need for collaboration between academia, research, and industry to handle large datasets and manage these methods.

1.2.3 Commonly Used Signals in Sleep Studies

Three commonly used signals in sleep studies are:

1. Electroencephalogram (EEG)

2. Electrooculogram (EOG)

3. Electromyogram (EMG)

1.2.4 Sleep Stages

The sleep cycle consists of three stages, including non-rapid eye movement (NREM) and rapid eye movement (REM), and is divided into:

- **Stage 1:** Wakefulness to light sleep with theta waves ($4-8$ Hz)
- **Stage 2:** Deeper sleep with K-complexes and frequencies ($0.5-2$ Hz)
- **Stage 3:** NREM-3 and NREM-4 with delta waves ($0.5-4$ Hz), synchronized with neuronal activity

1.2.5 EEG Data Processing

EEG data is stored in EDF format, often compressed in multiple channels. During data capture, EEG data is gathered from various brain locations to obtain spatial information. Data preprocessing includes algorithms for removing outliers and unnecessary data to enhance quality. This is followed by feature extraction and labeling according to AASM guidelines.

Chapter 2

Literature Review

1. The study shows challenges associated with arousal during sleep, which helps us understand the physiological effects during changes in sleep stages. It includes cognitive impairment and heart rate. To overcome these challenges, electroencephalography (EEG) signals data are used through machine learning (ML) algorithms. The study achieved an intensity level and sensitivity of 82.68%, specificity of 95.68%, and AUROC of 96.30% across the sleep stages [**?**].

2. This study highlights the important approach of ensemble learning for misclassified data in the form of time and frequency. It extracts feature data as per AASM guidelines and uses class-balanced random sampling techniques to improve model performance. The approach exhibits high accuracy and mean-F1 score, demonstrating robustness across individual sleep stages, irrespective of the efficiency or approach [**?**].

3. The study aims to classify optimal sleep stage classification using polysomno-

graphic signals. Despite the model's complexity, it captures interdependency and nonlinearity across all data types (EEG, EMG, EOG, and ECG). The model achieved an accuracy of 74%, which is notable given the minimal data used, and effectively differentiates between sleep stages [**?**].

4. This study addresses the automation of polysomnography, traditionally a manual task, using deep recurrent and convolutional neural networks (RCNN). The author collected PSG data and achieved accuracies of 87.6% for sleep staging, 88.2% for sleep apnea, and 84.7% for limb movements. This demonstrates the RCNN model's robustness and the potential impact of deep neural networks in mimicking human intervention [**?**].

5. The impact of computer advancements and automation in sleep stage scoring is examined. In the field of AI, various adaptive and parallel computing models, such as artificial neural networks (ANN), provide promising solutions. These models enable fast classification without compromising performance [**?**].

6. This study explores ML-based sleep stage scoring automation with time-frequency analysis. EEG data was used for feature extraction according to AASM standards. Ensemble learning was employed for sleep stage classification, focusing on class-balanced and random sampling. The study reports on mean F1-Score and accuracy, indicating practical applications [**?**].

7. The study discusses how feature extraction using ML techniques can improve over traditional methodologies by leveraging bio-electric signals. It explores how researchers identify sleep stages with quality assessment based

on the PICO framework. Despite advancements, a gap remains between ML models and signal processing. Further research should focus on robust technology for sleep staging automation, aiding clinical decision-making [**?**].

8. Although ML techniques are beneficial for automation, their adoption in daily clinical practice is limited. Automated and semi-automated deep learning techniques show promise, with increased computational power leading to notable results. Despite high accuracy in datasets, diverse datasets are required for deep learning techniques to achieve promising results [**?**].

Chapter 3

Machine Learning Techniques Implementation

3.1 Objective

The objective of this project is to develop a machine learning model that can accurately classify different sleep stages using EEG data. The primary focus is on preprocessing EEG data, handling class imbalance through resampling, and applying machine learning techniques to achieve high accuracy in classifying sleep stages.

3.2 Theory

Sleep is divided into multiple stages, each with distinct physiological characteristics. These stages can be identified using EEG (Electroencephalogram) signals, which measure electrical activity in the brain. The stages are typically divided into:

- Stage 1 (N1): Light sleep, transitioning between wakefulness and sleep.
- Stage 2 (N2): Light sleep, characterized by sleep spindles and K-complexes.
- Stage 3 (N3): Deep sleep, also known as slow-wave sleep.
- Stage 4 (N4): REM (Rapid Eye Movement) sleep, associated with dreaming.
- REM (Stage 5): Deep sleep with vivid dreaming.

Machine learning, specifically classification algorithms, can be employed to predict these stages based on the features extracted from EEG signals. Handling class imbalance, where some sleep stages are underrepresented, is crucial to ensure the model's performance.

3.3 Methodology

3.3.1 Data Preprocessing

Loading EEG Data The EEG data, consisting of raw EEG signals, was loaded and annotated. The data was then segmented into epochs of 30 seconds, corresponding to different sleep stages.

Epoch Creation Each epoch was labeled according to the sleep stage it belongs to. The data was then structured into a format suitable for machine learning, where each epoch is represented by its corresponding features (flattened EEG signal).

Class Imbalance Handling The data exhibited significant class imbalance, with some sleep stages being more prevalent than others. To address this, the SMOTE (Synthetic Minority Over-sampling Technique) algorithm was applied. SMOTE generates synthetic samples for underrepresented classes, balancing the dataset.

Model Training A RandomForestClassifier was used to train the model. The training process was conducted iteratively across several epochs, allowing the model to improve over time. The model was evaluated using various metrics, including accuracy, ROC-AUC, and log loss.

3.3.2 Model Development

Feature Engineering The EEG epochs were flattened into feature vectors, with each feature representing a point in the EEG signal over time.

Model Training A RandomForestClassifier was used for training the model. The training process was iterative, conducted across several epochs to allow continuous improvement. The model's performance was evaluated using various metrics, including accuracy, ROC-AUC, and log loss.

Random Forest Random forests, or random decision forests, are an ensemble learning method used for classification, regression, and other tasks. This method operates by constructing multiple decision trees during training. For classification tasks, the output of the random forest is the class chosen by the majority of the trees.

Evaluation Metrics

- **ROC-AUC Curve:** The ROC-AUC (Receiver Operating Characteristic -

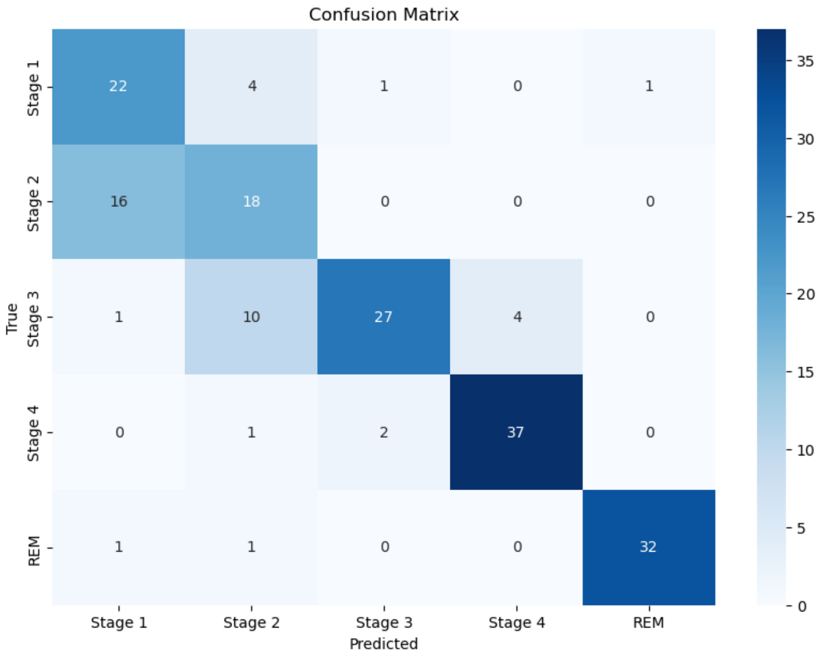

Figure 3.1. Confusion Matrix for Random Forest Classifier

Area Under Curve) curve was plotted to evaluate the model's capability to differentiate between different sleep stages.

- **Loss Curve:** The loss curve was plotted to observe the decrease in the model's log loss over the training epochs.
- **Accuracy Curve:** The accuracy curve was plotted to track the model's accuracy on both the training and testing datasets across multiple epochs.

3.3.3 Gradient Boosting

Gradient boosting is an ensemble learning method that builds a model in a stage-wise fashion from an ensemble of weak learners, typically decision trees. It works by iteratively improving the model's performance through the addition of new

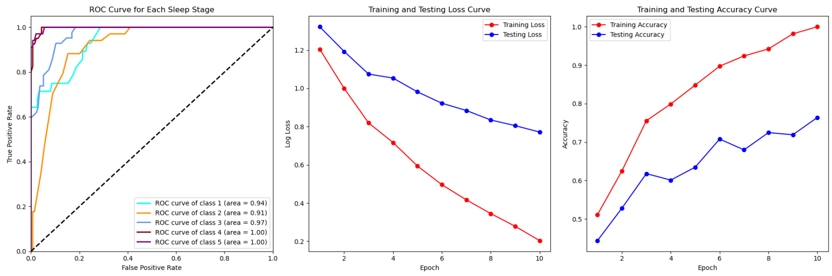

Figure 3.2. Accuracy and Loss Curves for Random Forest Classifier

trees that correct the errors made by the previous trees. For classification tasks, the final model is a weighted sum of all the weak learners, and the output is the class with the highest probability.

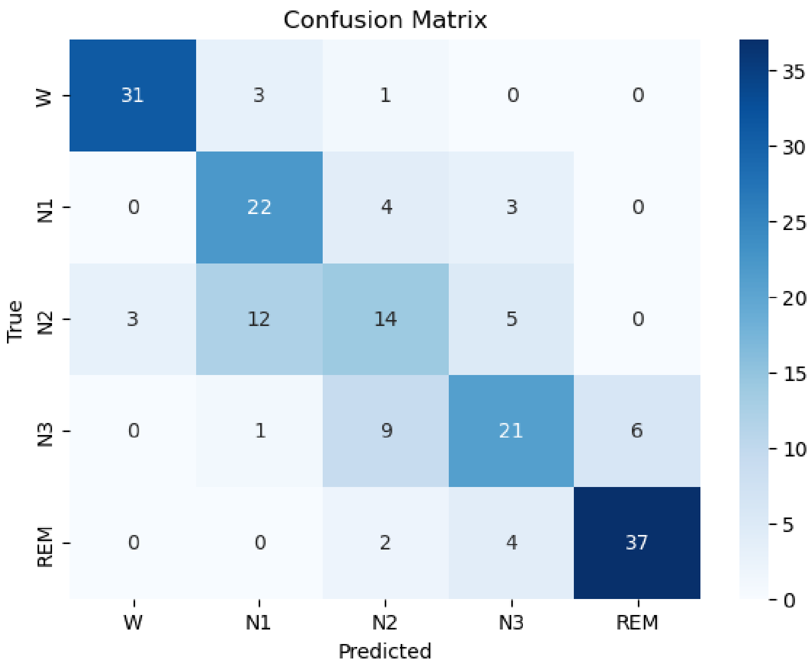

Figure 3.3. Confusion Matrix for Gradient Boosting Classifier

Evaluation Metrics

- **ROC-AUC Curve:** The ROC-AUC (Receiver Operating Characteristic - Area Under Curve) curve was plotted to evaluate the model's capability to differentiate between different sleep stages.

- **Loss Curve:** The loss curve was plotted to observe the decrease in the model's log loss over the training epochs.

- **Accuracy Curve:** The accuracy curve was plotted to track the model's accuracy on both the training and testing datasets across multiple epochs.

Figure 3.4. Accuracy and Loss Curves for Gradient Boosting Classifier

3.3.4 Combined Techniques: Gradient Boosting, Random Forest, and Voting Classifier

Gradient Boosting Classifier Gradient boosting is an ensemble learning method that builds a model incrementally from a set of weak learners, typically decision trees. Each new tree corrects the errors made by the previous trees, and the final model is a weighted sum of all the weak learners. This method enhances the model's predictive accuracy by focusing on the mistakes of earlier trees.

Random Forest Classifier Random forests are an ensemble learning method for classification and regression tasks that constructs multiple decision trees during training. Each tree is trained on a random subset of the data, and the final prediction is determined by the majority vote from all the trees. This method reduces overfitting and improves generalization by averaging the predictions of many trees.

Voting Classifier The Voting Classifier is an ensemble method that combines multiple models to improve classification performance. It aggregates the predictions from different classifiers (e.g., Gradient Boosting, Random Forest) and makes a final prediction based on the majority vote or average probability. This method leverages the strengths of various models to enhance overall predictive accuracy.

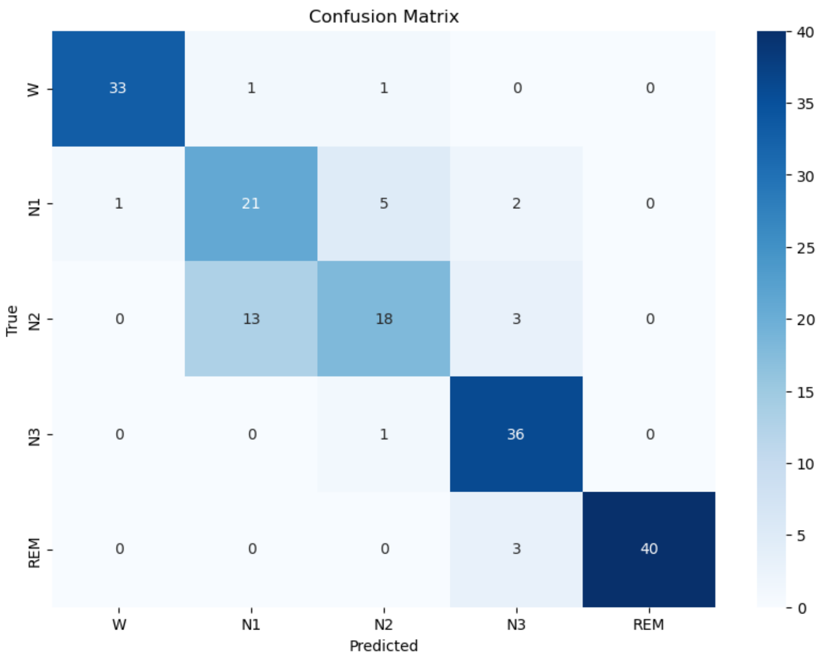

Figure 3.5. Confusion Matrix for Combined Techniques: Gradient Boosting, Random Forest, and Voting Classifier

Evaluation Metrics

- **ROC-AUC Curve:** The ROC-AUC (Receiver Operating Characteristic - Area Under Curve) curve was plotted to evaluate the combined models' ability to distinguish between different sleep stages.

- **Loss Curve:** The loss curve was plotted to observe how the log loss decreased over the training epochs for the combined techniques.

- **Accuracy Curve:** The accuracy curve was plotted to track the accuracy of the combined models on both the training and testing datasets across multiple epochs.

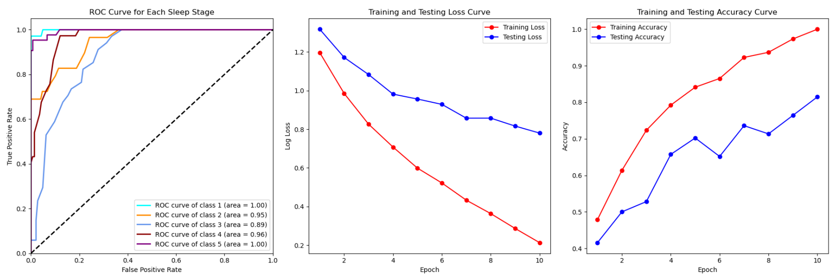

Figure 3.6. Accuracy and Loss Curves for Combined Techniques: Gradient Boosting, Random Forest, and Voting Classifier

3.3.5 Bagging Classifier

Bagging Classifier Bagging, or Bootstrap Aggregating, is an ensemble learning technique that improves the stability and accuracy of machine learning algorithms. It works by training multiple instances of the same learning algorithm on different subsets of the training data, which are generated through bootstrapping (random sampling with replacement). Each model makes predictions, and the final output is determined by aggregating the predictions from all individual models, typically through voting for classification tasks or averaging for regression tasks. This method reduces variance and helps in mitigating overfitting.

Evaluation Metrics

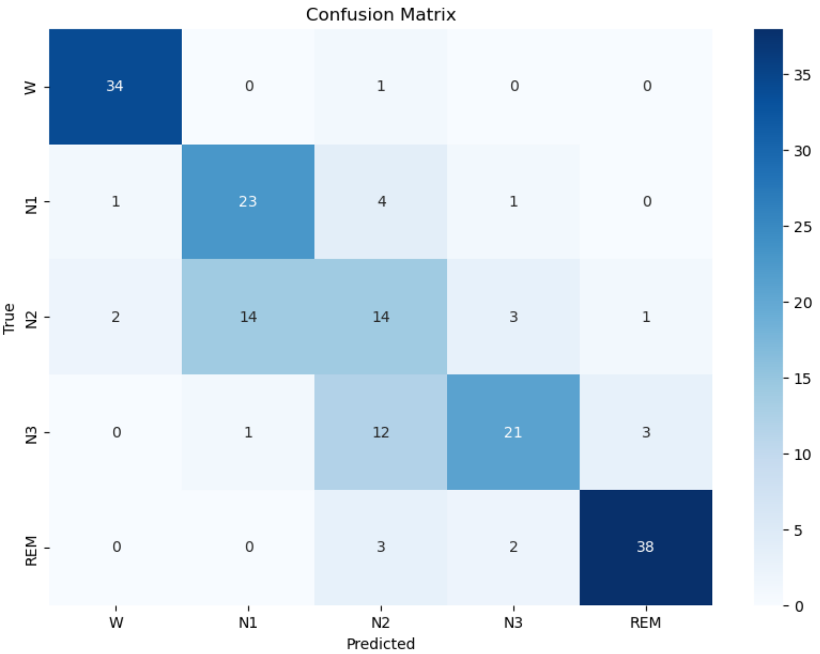

Figure 3.7. Confusion Matrix for Bagging Classifier

- **ROC-AUC Curve:** The ROC-AUC (Receiver Operating Characteristic - Area Under Curve) curve was plotted to assess the Bagging Classifier's ability to distinguish between different sleep stages.

- **Loss Curve:** The loss curve was plotted to observe the Bagging Classifier's log loss reduction over the training epochs.

- **Accuracy Curve:** The accuracy curve was plotted to monitor the Bagging Classifier's accuracy on both training and testing datasets across multiple epochs.

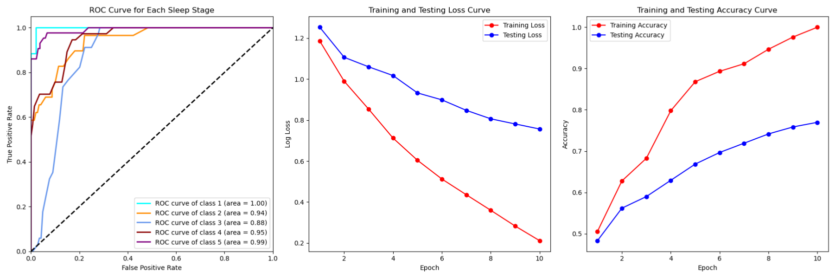

Figure 3.8. Accuracy and Loss Curves for Bagging Classifier

Chapter 4

Results and Discussion

4.1 Classifier Performance

The following table summarizes the performance of different classifiers:

TABLE 4.1
Performance Metrics for Various Classifiers

Classifier	Accuracy	Precision	Sensitivity	F-Measure
Random Forest	0.775281	0.779582	0.780000	0.772726
Gradient Boosting	0.713483	0.698613	0.708304	0.700800
CLF	0.797753	0.787969	0.782399	0.781475
Bagging Classifier	0.775281	0.759451	0.766824	0.755140

4.2 Classification Report

The classification report for the model is as follows:

TABLE 4.2
Classification Report

Class	Precision	Recall	F1-Score	Support
W	0.8974	1.0000	0.9459	35
N1	0.6286	0.7586	0.6875	29
N2	0.6667	0.4118	0.5091	34
N3	0.7568	0.7568	0.7568	37
REM	0.8478	0.9070	0.8764	43
Accuracy		0.7753		178
Macro Avg	0.7595	0.7668	0.7551	178

4.3 Performance Plots

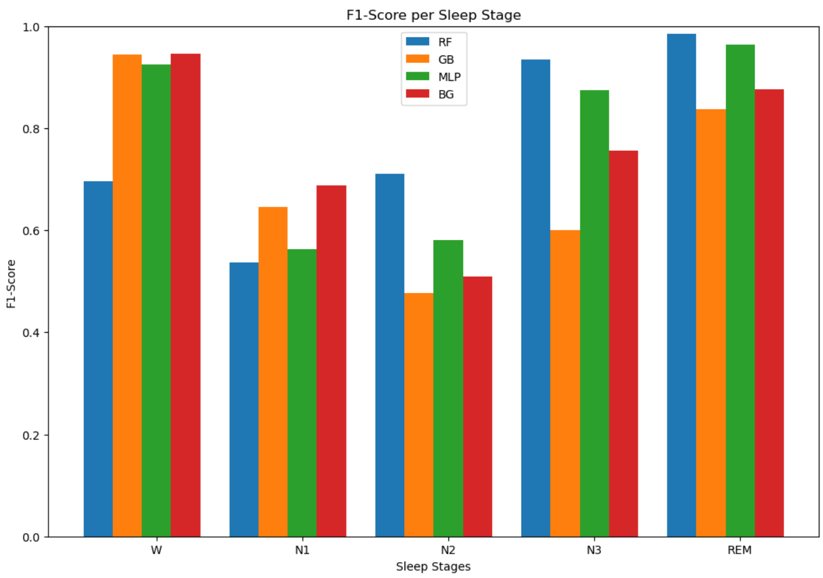

Figure 4.1. F1 Score Plot

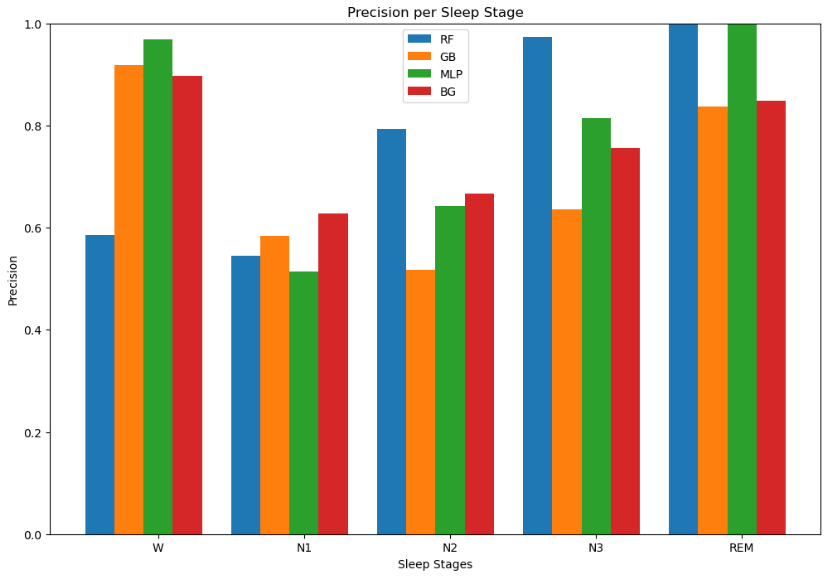

Figure 4.2. Precision Score Plot

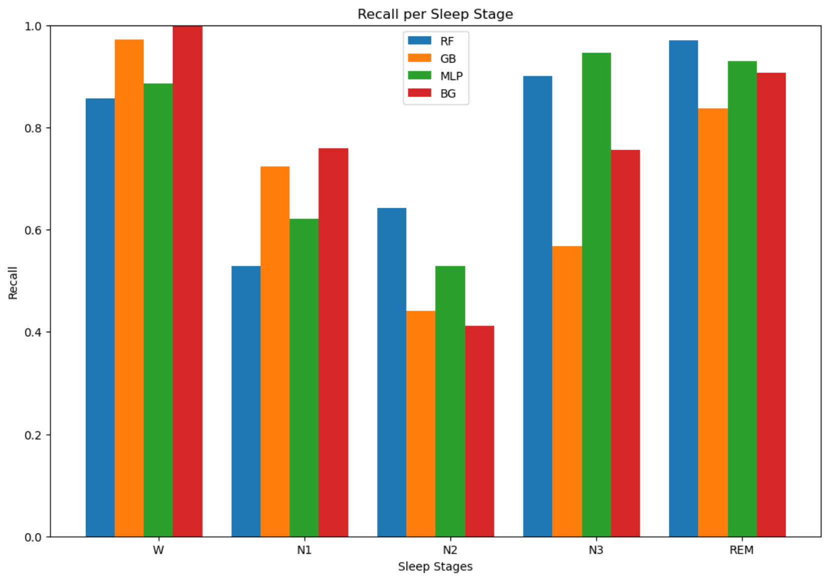

Figure 4.3. Recall Score Plot

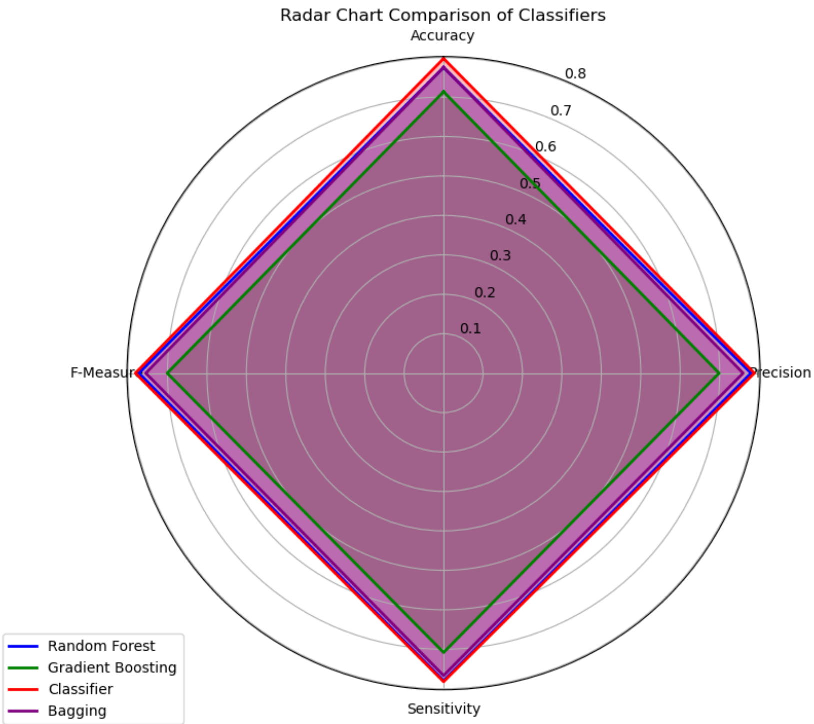

Figure 4.4. Radar Chart of Classification Metrics

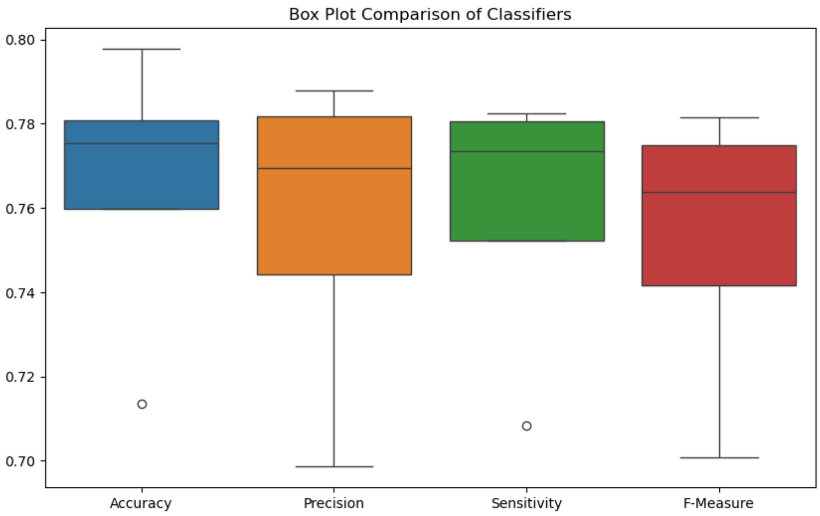

Figure 4.5. Box Plot of Model Performance Metrics

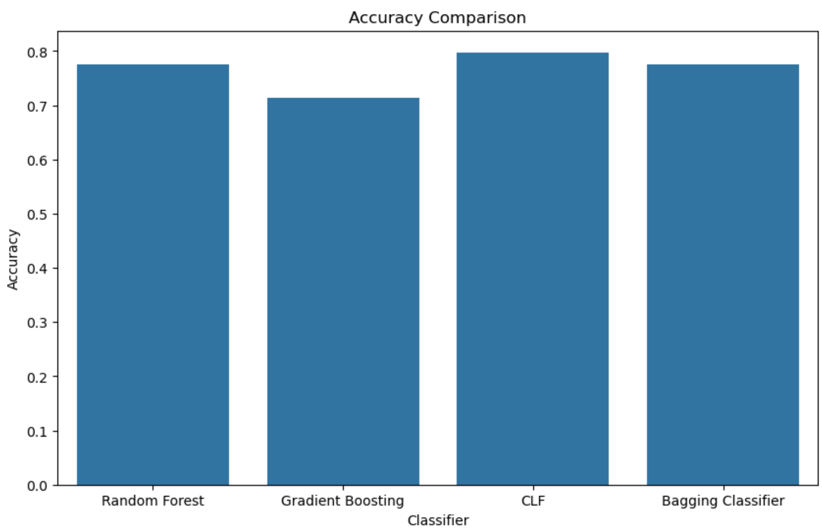

Figure 4.6. Accuracy Comparison Bar Plot for Different Classifiers

Chapter 5

Conclusions

5.1 Conclusion

The study demonstrates the effectiveness of various machine learning techniques in classifying sleep stages using EEG data. The RandomForestClassifier, GradientBoostingClassifier, and Bagging Classifier showed promising results, with the combined techniques providing enhanced accuracy and robustness. The use of resampling methods to address class imbalance was crucial in improving model performance. Future work could explore deep learning methods and additional feature extraction techniques to further enhance classification accuracy.

Chapter 6

Reference

Bibliography

[1] Hyewon Han *et al.*, "Classification and automatic scoring of arousal intensity during sleep stages using machine learning," *Scientific Reports*, vol. 14, no. 1, Dec. 2024. DOI: 10.1038/s41598-023-50653-9.

[2] Orestis Tsinalis *et al.*, "Automatic Sleep Stage Scoring Using Time-Frequency Analysis and Stacked Sparse Autoencoders," *Annals of Biomedical Engineering*, vol. 44, no. 5, May 2016. DOI: 10.1007/s10439-015-1444-y.

[3] Anna Krakovská and Kristína Mezeiová, "Automatic sleep scoring: A search for an optimal combination of measures," *Artificial Intelligence in Medicine*, vol. 53, no. 1, Sep. 2011. DOI: 10.1016/j.artmed.2011.06.004.

[4] Siddharth Biswal *et al.*, "Expert-level sleep scoring with deep neural networks," *Journal of the American Medical Informatics Association*, vol. 25, no. 12, Dec. 2018. DOI: 10.1093/jamia/ocy131.

[5] Marina Ronzhina *et al.*, "Sleep scoring using artificial neural networks," *Sleep Medicine Reviews*, vol. 16, no. 3, Jun. 2012. DOI: 10.1016/j.smrv.2011.06.003.

[6] Hadeel Alsolai *et al.*, "A Systematic Review of Literature on Automated Sleep

Scoring," *IEEE Access*, vol. 10, pp. 79419-79443, 2022. DOI: 10.1109/ACCESS.2022.3194145.

[7] Luigi Fiorillo *et al.*, "Automated sleep scoring: A review of the latest approaches," *Sleep Medicine Reviews*, vol. 48, Dec. 2019. DOI: 10.1016/j.smrv.2019.07.007.

[8] Oliver Faust *et al.*, "A review of automated sleep stage scoring based on physiological signals for the new millennia," *Computer Methods and Programs in Biomedicine*, vol. 176, Jul. 2019. DOI: 10.1016/j.cmpb.2019.04.032.

[9] Panteleimon Chriskos *et al.*, "A review on current trends in automatic sleep staging through bio-signal recordings and future challenges," *Sleep Medicine Reviews*, vol. 55, Feb. 2021. DOI: 10.1016/j.smrv.2020.101377.