ADHD PREDICTION USING MACHINE LEARNING ALGORITHMS

BY

SADNAM SANIAT
ID: 192-15-2794

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised By

SHAH MD TANVIR SIDDIQUEE
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

NARAYAN RANJAN CHAKRABORTY
Associate Professor and Associate Head
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "ADHD prediction using machine learning algorithms" submitted by **Sadnam Saniat** 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 14th July, 2024.

BOARD OF EXAMINERS

Dr. Sheak Rashed Haider Noori — **Board Chairman**
Professor & Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Most. Hasna Hena — **Internal Examiner 1**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Md. Ferdouse Ahmed Foysal — **Internal Examiner 2**
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

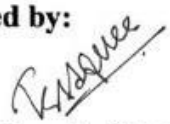
Dr. Md. Arshad Ali — **External Examiner**
Professor
Department of the CSE
Hajee Mohammad Danesh Science and Technology University

©Daffodil International University, Bangladesh

DECLARATION

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

Supervised by:

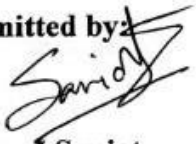


Shah Md Tanvir Siddiquee
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Narayan Ranjan Chakraborty
Associate Professor and Associate Head
Department of CSE
Daffodil International University

Submitted by:



Sadnam Saniat
ID: -192-15-2794
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project successfully.

I am grateful and wish our profound indebtedness **Shah Md Tanvir Siddiquee, Assistant Professor,** Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine learning*" to carry out this project. Her 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.

I would like to express our heartiest gratitude to the **Dr. Sheak Rashed Haider Noori, Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

I would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, I must acknowledge with due respect the constant support and patience of my parents.

ABSTRACT

This study analyzes the use of machine learning algorithms to predict Attention Deficit Hyperactivity Disorder (ADHD) based on a number of feature sets. A dataset from online containing demographic information, personality characteristics, and medical indicators was used. Among the classifiers used, the logistic regression model achieved the highest accuracy of 98.17%. Data preprocessing consisted of cleaning, transforming, and encoding features, followed by feature selection and exploratory data analysis. The logistic regression model, trained on the preprocessed dataset, outperformed other classifiers including random forest, voting, gradient boosting, Gaussian Naive Bayes, and decision tree models, with an accuracy of 98.17%. The model showed high precision, recall, and F1-score, indicating that it was effective in distinguishing between people with and without ADHD. The dataset's attributes were important in identifying key predictors of ADHD, providing helpful information for clinical decision-making. Ethical concerns about data privacy and algorithmic bias were addressed all through the study to ensure responsible implementation. Overall, this study shows the potential of machine learning to improve ADHD diagnosis and shows the importance of comprehensive data collection and rigorous model evaluation in healthcare applications.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	ii
Declaration	iii
Acknowledgements	iv
Abstract	v
Table of contents	vi-ix
List of Figures	x
List of Tables	xi
CHAPTER	
CHAPTER 1: INTRODUCTION	**1-4**
1.1 Overview	1
1.2 Background and Present State	2
1.3 Problem Statement	2
1.4 Objectives	2
1.5 Scope and Limitations	3
1.6 Report Organization	3
1.7 Summary	4

CHAPTER 2: LITERATURE REVIEW **5-11**

2.1 Overview 5

2.2 Related Works 5

2.3 Comparison between existing works 9

2.4 Open Issues 9

2.5 Summary 11

CHAPTER 3: METHODOLOGY/ REQUIREMENT ANALYSIS & DESIGN SPECIFICATION **12-21**

3.1 Overview 12

3.2 Proposed Methodology/ System Design 12

3.3 Hardware/ Software Requirement 19

3.4 Project Management and Financial Analysis 19

3.5 Summary 20

CHAPTER 4: IMPLEMENTATION **22-24**

4.1 Overview 22

4.2 Train Model/ Prototype Design 22

4.3 System Testing/ Model Evaluation 23

4.4 Summary 23

CHAPTER 5: RESULT AND ANALYSIS **25-34**

5.1 Overview 25

5.2 Experimental/ Simulation Result 25

5.3 Performance/ Comparative Analysis 33

5.4 Summary 34

CHAPTER 6: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY **35-37**

6.1 Impact on Life 35

6.2 Impact on Society & Environment 35

6.3 Ethical Aspects 36

6.4 Sustainability Plan 36

6.5 Summery 37

CHAPTER 7: CONCLUSION AND FUTURE WORK **38-39**

7.1 Conclusions 38

7.2 Future Suggested Works 38

7.3 Limitations/ Conflict of Interest 39

REFERENCES **40-42**

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.1: Proposed LR Architecture	14
Figure 3.2: Proposed GB Architecture	14
Figure 3.3: Proposed DT Architecture	15
Figure 3.4: Proposed RF Architecture	16
Figure 3.5: Proposed GNB Architecture	17
Figure 3.6: Proposed Voting Classifier Architecture	17
Figure 3.7: Work Flow	18
Figure 5.1: Confusion Matrix LR	27
Figure 5.2: Confusion Matrix GNB	28
Figure 5.3: Confusion Matrix DT	29
Figure 5.4: Confusion Matrix GBC	30
Figure 5.5: Confusion Matrix RF	31
Figure 5.6: Confusion Matrix Voting Classifier	32
Figure 5.7: Accuracy Graph	34

LIST OF TABLES

TABLES	PAGE NO
Table 2.3. Comparative Analysis Table with Previous Work	9
Table 3.4. Project Management Table	20
Table 3.5: Estimated Cost	20
Table 5.1: Performance Evaluation	33

CHAPTER 1
INTRODUCTION

1.1 Overview

Attention Deficit Hyperactivity Disorder (ADHD) is a neurodevelopmental disorder defined by persistent patterns of inattention, hyperactivity, and impulsivity that severely impair daily functioning in a variety of settings. ADHD is a complex condition that is frequently identified through comprehensive clinical assessments such as interviews, behavioral observations, and standardized rating scales. However, advances in machine learning (ML) techniques provide promising opportunities to improve ADHD diagnosis and prediction through data-driven approaches J.-W. Kim et al [1]. The purpose of this study is to look into the use of machine learning algorithms for predicting ADHD using a diverse set of features collected from individuals. The dataset, obtained from Kaggle, includes demographic data, behavioral traits, medical history, and other relevant factors that could impact ADHD diagnosis. This study uses a structured methodology that includes data preprocessing, feature selection, model training, and evaluation to develop robust predictive models capable of identifying individuals at risk of ADHD Navya Sethu and R. Vyas et al. [2]. The significance of this study comes from its potential for adding to traditional diagnostic approaches with knowledge based on data, thereby increasing the efficiency and accuracy of ADHD detection. ML algorithms can analyze large amounts of heterogeneous data, uncover hidden patterns, and make personalized predictions based on individual characteristics. Furthermore, the use of ML in ADHD prediction shows promise for early intervention, treatment planning, and resource allocation in clinical settings. Y. Zhang-James et al. [3]. Through this research, we hope to add to the increasing amount of literature on the connection of mental health and machine learning by providing valuable insights into the predictive capabilities of ML algorithms for ADHD diagnosis. Finally, the study's findings may inform the development of decision support systems and tools that can help healthcare professionals in the early detection and management of ADHD, thereby improving outcomes for those affected by the condition.

1.2 Background and Present State

Before looking into the main content of this report, it is important to have an accurate grasp of the key concepts and terms relevant to the study. This section will define and explain terms related to Attention Deficit Hyperactivity Disorder (ADHD), machine learning algorithms, and predictive modeling techniques. Furthermore, it will discuss the importance of ADHD diagnosis, the potential benefits of machine learning in healthcare, and the rationale for studying the intersection of these domains. By laying out these preliminaries, readers will have the necessary background knowledge to engage with the following chapters and understand the context and implications of the research presented herein.

1.3 Problem Statement

The rationale for conducting this study comes from the urgent need to address the challenges related to diagnosing and managing hyperactivity disorder. Despite its increasing prevalence and significant impact on people's lives, ADHD diagnosis remains complex and subjective, resulting in variability in evaluations and prevents appropriate intervention. Also hope to improve the efficiency and accuracy of ADHD prediction by utilizing machine learning (ML) algorithms that have shown promise in analyzing complex datasets and identifying patterns. This study is designed to fill a gap in existing diagnostic approaches by using machine learning techniques to integrate diverse sources of data, such as demographic information, behavioral traits, and medical history, to create robust predictive models. Finally, the use of machine learning in ADHD diagnosis has the potential to improve early detection, personalized treatment planning, and outcomes for people with this neurodevelopmental disorder.

1.4 Objective

The primary purpose by the urgent need to improve the accuracy, efficiency, and accessibility of ADHD diagnosis and intervention. ADHD is a common neurodevelopmental disorder that affects people throughout their lives, with serious implications for academic, social, and professional success. However, diagnosing ADHD can be challenging due to its variable presentation, overlap with other mental health

conditions, and reliance on subjective assessments. Traditional diagnostic methods frequently rely on subjective judgments, resulting in variability in diagnoses and delays in receiving appropriate care. We hope to address these challenges by developing predictive models that can use diverse datasets to identify patterns associated with ADHD. We hope to address these challenges by developing predictive models that can use diverse datasets to identify patterns associated with ADHD. Such models have the potential to improve diagnostic accuracy, early intervention, and outcomes for people with ADHD. Furthermore, by using online available data sources such as Kaggle, we can speed up the development and deployment of ML-based tools for ADHD prediction, making them more accessible to doctors, researchers, and individuals seeking assistance. Finally, our motivation comes from the belief that including machine learning techniques into ADHD diagnosis can result in more personalized, efficient, and equitable mental healthcare practices.

1.5 Scope and Limitations

The scope of this study is the development of robust machine learning models capable of accurately predicting ADHD diagnosis based on a variety of features. These models are expected to have excellent predictive performance, as measured by metrics like accuracy, precision, recall, and area under the receiver operating characteristic curve (AUC-ROC). Furthermore, the study aims to identify key demographic factors, behavioral traits, and medical indicators that have the greatest impact on ADHD prediction, providing valuable insights into the disorder's underlying features. Furthermore, the study aims to compare the performance of various machine learning algorithms and feature selection techniques, providing guidance on the most effective approaches to ADHD prediction. Ultimately, the findings of this study are expected to inform the development of decision support tools and interventions to aid in early ADHD detection, personalized treatment planning, and improved outcomes for individuals suffering from this neurodevelopmental disorder.

1.6 Report Organization

This report aims to provide a thorough overview of the research on ADHD prediction using machine learning algorithms. Introduction chapter provides an overview of the study's

objectives, meaning, and scope, developing the background for the research on ADHD prediction using machine learning. Background study chapter reviews the existing literature and theoretical foundations pertaining to ADHD diagnosis, machine learning techniques, and predictive modeling approaches. Relevant literature, theories, and previous research on skin disease classification, deep and transfer learning techniques. Methodology describes the study's methodology, which included data collection, preprocessing, feature selection, model training, and evaluation. The study's methodology, which includes data collection, preprocessing, model selection, training, and evaluation procedures. Experimental Results presents the study's findings and outcomes, such as the performance of machine learning models in predicting ADHD and the insights gained from the analysis. The Impact on Society discusses the research findings' potential societal implications, such as those for healthcare practices, early intervention, and public health policy. Conclusion and Implications for Future Research chapter summarizes the key findings, draws conclusions from the study, makes recommendations for future research directions, and discusses the broader implications of the results.

1.7 Summary

In the beginning, the Introduction chapter explains the context as well as the purpose of the research, stressing the importance of predicting ADHD using ML algorithms. It provides the basis that led to the investigation of applying data science in improving the diagnosing and treatment of ADHD. The chapter also provides the research questions and the scope of the study while considering its applicability to the improvement of clinical practices and patients' lives. Also, it mentions the reasons that justified the choice of certain methodologies of machine learning and key concerns for ethic in the implementation context. In sum, the Introduction establishes a foundation for an extensive review of the use and prediction of probabilistic ADHD advancement through what the author deems as advanced algorithms.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

The Literature Review chapter therefore focuses on analyzing literature in relation to ADHD diagnosis and the use of machine in healthcare. It covers different sorts of previously applied machine learning techniques for the early identification of mental health disorders and the advantages and disadvantages of each. In this regard, the various review methods involve making key annotations of research studies, their methodologies, findings, and discrepancies. Besides, it covers the pros and cons of using machine learning models in clinical practice and its moral implications. This chapter gives the relevant background information that is important in the establishment of the proposed methodology for ADHD prediction.

2.2 Related Works

Literature Review looks into a variety of studies regarding ADHD diagnosis and the utilization of machine learning in the healthcare sector. It covers different sorts of previously applied machine learning techniques for the early identification of mental health disorders and the advantages and disadvantages of each. It can also summarize important studies which have been done in the particular field of study mentioning the methods used, the conclusions drawn and the issues left unexplored. Moreover, it discusses some of the moral implications and dilemmas related to the use of machine learning algorithms in supporting clinical practices. This chapter gives the relevant background information that is important in the establishment of the proposed methodology for ADHD prediction. M. L.Cervantes-Henríquez et al.[5] used, DSM-IV symptomatology's latent class cluster analysis was used to determine the severity of ADHD. SNPs in DRD4, SNAP25, and ADGRL3 were linked to ADHD severity across various domains, according to family-based association testing. Genetic and demographic data were used by machine learning algorithms to build predictive models of the severity of ADHD. The models demonstrated

possible pleiotropic effects of certain SNPs on ADHD severity by efficiently differentiating between non-severe and severe ADHD across particular symptom domains.

A. Kautzky et al. [6] conducted a PET investigation where the serotonin transporter (SERT) binding potential was measured with CDASB in 16 ADHD patients and 22 healthy controls.. The HTR1A, HTR1B, HTR2A, and TPH2 genes were genotyped at thirty SNPs. Key areas of interest (ROIs) and discriminative SNPs were found using random forest (RF) machine learning, which also produced a mean validation accuracy of 0.82 with a balanced sensitivity and specificity. The model highlights the importance of the SERT, HTR1B, and HTR2A genes in ADHD, highlighting the necessity for accurate diagnostic instruments as well as possible disease-specific consequences. Md. Maniruzzaman et al. [7] proposed an EEG-based machine learning (ML) algorithm to distinguish ADHD from healthy children. Data from 61 ADHD and 60 healthy children aged 7–12 were analyzed. Morphological and time-domain EEG features were extracted and refined using t-tests and LASSO. Four ML algorithms were applied, with LASSO-SVM achieving the highest accuracy (94.2%), sensitivity (93.3%), and AUC (0.964). MLP was the next best classifier. The combination of LASSO and SVM offers promise for early ADHD diagnosis. M Duda et al. [8] used six machine learning models were trained and validated using forward feature selection, undersampling, and 10-fold cross-validation using 65-item Social Responsiveness Scale scores from 2925 persons with ASD (n = 2775) or ADHD (n = 150). With an AUC of 0.965, five behaviors from this screening test were able to identify between ASD and ADHD with accuracy. These results underscore the possibility for effective initial risk assessment and screening and validate the usefulness of machine learning in identifying autism and ADHD. Anshu Parashar et al. [9] proposed A machine learning model for categorizing children with ADHD and healthy controls is developed in this study using EEG signals. Classifiers such as AdaBoost, Random Forest, and Support Vector Machine were used to assess EEG data from 60 children with ADHD and 60 healthy controls. Considering every channel in the Right Hemisphere, the AdaBoost classifier had the highest accuracy (84%), while maintaining a 96% sensitivity. The efficacy of EEG signals in diagnosing ADHD is demonstrated by this methodology. Xiaolong Peng et al. [10] investigated 55 ADHD subjects' and 55 healthy controls' high-resolution MR pictures. 340 cortical characteristics were produced by the FreeSurfer program from 68 brain regions.

The best features for classifying ADHD were chosen using the F-score and SFS techniques. ELM outperformed SVM-Linear (84.73%) and SVM-RBF (86.55%) with an accuracy of 90.18% using eleven combined features. With different sample sizes, ELM demonstrated better computational robustness and efficiency. Frontal, temporal, occipital, and insular lobes showed distinct variations between patients with ADHD and healthy individuals.

D. Andrea et al. [11] used a variety of characteristics and symptom data to identify ADHD using machine learning techniques. Numerous machine learning methods, such as logistic regression, k-nearest neighbors, SVMs, Random Forest, XGBoost, and ANN, were used with the CAP dataset. With an AUC of 0.99, the ANN model had the highest accuracy. These results demonstrate machine learning's effectiveness in comprehending the etiological elements of ADHD and point to the technology's potential to improve diagnostic accuracy. You can get the code by visiting the GitHub Repository. Hanna Christiansen et al. [12] used the 26 items of the Conners' Adult ADHD Rating Scales (CAARS-S: S), individuals with ADHD, obesity, problematic gambling, and a control group were distinguished with a worldwide accuracy of 0.80 using the LightGBM algorithm in machine learning. Across diseases, recall varied from 0.58 to 0.87 and precision from 0.78 to 0.92. The best 5 and best 10 item models produced less good fits, suggesting that CAARS-S may be useful for multiclassifying conditions similar to ADHD. William Das et al. [13] suggests a solid machine learning framework based on pupil-size dynamics for automated detection of ADHD. The support vector machine attained an average AUROC of 85.6% with 77.3% sensitivity and 75.3% specificity by integrating binary classification techniques, visualization, and pupillometric feature engineering. By using ten-fold nested cross-validation to a dataset comprising fifty patients, 218 noteworthy features were found, providing insight into the correlation between pupil-size dynamics and ADHD. Robust AUROC values highlight the promise of pupillometrics as a discriminative biomarker for the identification of ADHD, even with a small sample size. This groundbreaking work demonstrates how oculometric paradigms and machine learning can be used to diagnose ADHD. Akira Yasumura et al. [14] used fresh brain function data, machine learning—more especially, a support vector machine, or SVM—was used to predict the severity of the condition. Data, including behavioral, physiological, and age indicators, were gathered from 108 children with ADHD and 108 typically developing kids across several centers

88.71% sensitivity, 83.78% specificity, and 86.25% overall discrimination rate were attained by the SVM. Md. Maniruzzaman et al. [15] determine important risk variables for ADHD in children and to provide a categorization method based on machine learning. 45,779 children between the ages of three and seventeen had their data taken from the 2018–2019 National Survey of Children's Health. Eight machine learning classifiers were used for prediction, and relevant predictors for ADHD were retrieved using logistic regression. With an accuracy of 85.5%, sensitivity of 84.4%, specificity of 86.4%, and an AUC of 0.94, the RF classifier outperformed the LR with RF-based system in classifying ADHD cases. Anuradha et al. [16] introduces the SVM method for diagnosing ADHD, aiming to mitigate diagnostic complexity. This innovative attempt to diagnose ADHD using SVM is further enhanced by a genetic algorithm-based feature selection approach. Genetic algorithms offer solutions to complex issues. Pavol Mikolaset al. [17] used a linear SVM classifier trained on clinical records data (N=299 participants) to address the challenging diagnostic process of ADHD and identify the disorder in people with a range of mental illnesses. When it came to accurately identifying children and adolescents with ADHD, the SVM scored 66.1%. Nineteen features were combined for best performance, according to automated feature selection. The automatic detection of ADHD without compromising performance was made possible by real-world clinical data, underscoring the significance of taking into account all symptoms when making a diagnosis. WilliamDas et al. [18] used a linear SVM classifier trained on clinical records data (N=299 participants) to address the challenging diagnostic process of ADHD and identify the disorder in people with a range of mental illnesses. When it came to accurately identifying children and adolescents with ADHD, the SVM scored 66.1%. Nineteen features were combined for best performance, according to automated feature selection. The automatic detection of ADHD without compromising performance was made possible by real-world clinical data, underscoring the significance of taking into account all symptoms when making a diagnosis. Hui Wen Loh et al. [19] categorized studies on ADHD diagnosis using machine learning and deep learning techniques according to the diagnostic instrument used. Research shortcomings include a lack of attention to wearable device data use and a dearth of publicly available datasets beyond MRI. The review intends to stimulate more research to generate larger datasets and investigate other modalities for diagnosing ADHD, such as

motion data and ECG. In the end, it sees AI assisting a thorough clinical decision-making process for the diagnosis and follow-up of ADHD. Yanli Zhang-James et al. [20] conduct a study examining neuroanatomical alterations associated with ADHD in both children and adults using deep learning neural network classification models. In all age groups, structural MRI data successfully distinguish individuals with ADHD from control subjects, with children exhibiting superior prediction performance and effect sizes. The model's accurate prediction of ADHD in children, trained on adult data, suggests structural similarities between childhood and adult ADHD. This innovative method supports the pathophysiological continuity of ADHD across developmental stages.

2.3 Comparison between existing works

The comparative analysis assesses the performance of various machine learning algorithms in predicting ADHD. Logistic regression had the highest accuracy of 98.17%, outperforming other classifiers like random forest, voting, gradient boosting, Gaussian Naive Bayes, and decision tree models. The logistic regression model also had high precision, recall, and F1-score, indicating that it was effective in distinguishing between ADHD and non-ADHD individuals. These findings highlight the potential of logistic regression as a strong predictive model for ADHD diagnosis.

TABLE 2.3 COMPARATIVE ANALYSIS WITH PREVIOUS WORK

SL	Author Name	Used Algorithm	Best Accuracy with Algorithm
1.	J.-W.Kim et al. [1]	ML	SVM= 84.6%
2	Y. Zhang-James et al. [3]	ML & RNN	RF= 73%
3	J.Downs et al [4]	ML	LR=86%
4	Md. Maniruzzaman et al. [7]	ML	SVM=94.2%
5	Anshu Parashar et al. [9]	ML	Adaboost =84%
6	Xiaolong Peng et al. [10]	ML	SVM= 90.18%
10	Proposed Approach	ML	LR= 98.17%

2.4 Open Issues

The scope of this research includes the development and evaluation of machine learning (ML) models for predicting Attention Deficit Hyperactivity Disorder (ADHD) using a variety of feature sets. The study's specific goal is to look into the viability and efficacy of using machine learning algorithms to analyze demographic data, behavioral traits, and medical indicators to diagnose ADHD. The study will look into various machine learning techniques, such as logistic regression, decision trees, random forests, and gradient boosting, to determine the most accurate and reliable predictors of ADHD. Furthermore, the study will look into the potential benefits of ML-based ADHD prediction for early detection, personalized treatment planning, and better clinical outcomes. However, it is important to note that this study does not seek to replace traditional diagnostic methods, but rather to add to them with data-driven insights and predictive capabilities.

2.5 Summary

The following summary are inherent in the development and implementation of machine learning (ML) models for predicting Attention Deficit Hyperactivity Disorder (ADHD). Firstly, acquiring high-quality, diverse datasets with sufficient sample sizes and comprehensive feature sets poses a significant challenge due to the complexity and heterogeneity of ADHD. Additionally, ensuring the reliability and validity of ADHD diagnoses in the dataset is crucial for model accuracy. Furthermore, interpreting and explaining the decisions made by ML models in the context of ADHD diagnosis presents challenges, particularly in maintaining transparency and accountability. Finally, addressing ethical considerations, such as data privacy, algorithmic bias, and the potential impact on individuals' well-being, requires careful attention throughout the research process. Overcoming these challenges will be essential for the successful development and deployment of ML-based ADHD prediction tools in clinical practice.

CHAPTER 3
METHODOLOGY

3.1 Overview

The study focuses on predicting ADHD using machine learning algorithms. The study's specific goal is to create predictive models that can identify people at risk of ADHD based on statistics, behavioral characteristics, and medical indicators. This study's instrumentation includes a variety of data sources, including standardized evaluation methods, medical records, demographic surveys, and behavioral observations. The data will be examined and understood using machine learning techniques such as logistic regression, decision trees, random forests, and gradient boosting. Model development and evaluation will also involve the use of software tools and programming languages such as Python, as well as libraries such as scikit-learn and TensorFlow. The research topic and instrumentation have been designed to address the complexities of ADHD diagnosis while using advanced computational methods to improve predictive accuracy and clinical utility.

3.2 Proposed Methodology

In this proposed methodology, we are hoping to use machine learning algorithms to predict Attention Deficit Hyperactivity Disorder (ADHD) based on a variety of features. We aim to develop accurate and reliable predictive models to supplement traditional diagnostic approaches by following a structured approach that includes data selection, preprocessing, model training, and evaluation.

Data Collection:

Collect a suitable dataset from Kaggle that includes features relevant to ADHD prediction, such as statistics, behavioral traits, medical history, and ADHD diagnosis.

Data Pre- Processing:

Data Cleaning: Addresses missing values, outliers, and mistakes in the dataset.

Transform: Make necessary changes to the data, such as increasing or normalization.

Binary Labeling: Define the target variable for ADHD prediction and assign binary labels (0 for non-ADHD and 1 for ADHD).

Feature Selection:

Identify relevant features that help predict ADHD using techniques such as correlation analysis, feature importance, or domain knowledge.

Encoding:

Using techniques such as one-hot encoding or label encoding, encode categorical variables into numerical representations that machine learning algorithms can handle.

Exploratory Data Analysis (EDA):

Explore the dataset with graphs and methods of statistics to learn about feature distribution and variable relationships.

Model Training:

Divide the preprocessed data into training and testing sets. Use the training data to train machine learning models such as logistic regression, decision trees, random forests, and gradient boosting. In Given below I am shortly describing those models:

Logistic Regression

Logistic regression is a fundamental statistical model for binary classification that is widely utilized. In contrast to linear regression, it represents the relationship between the input variables and the likelihood of belonging to a specific class. The logistic function (sigmoid) is used in logistic regression to transfer the linear combination of input features to a probability score. This probabilistic technique enables intuitive interpretation as well as the estimation of class probabilities. It is noise-resistant and computationally efficient, and it works well with both continuous and categorical predictors.

Linear Combination:

$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$

Here, z represents the linear combination of input features, β_0 is the intercept term, β_1, β_2, ..., β_p are the coefficients for each input feature (x_1, x_2, ..., x_p).

Sigmoid Transformation:

$p = 1 / (1 + e^{(-z)})$

The sigmoid function (1 / (1 + e^(-z))) maps the linear combination (z) to a probability

score (p) between 0 and 1.

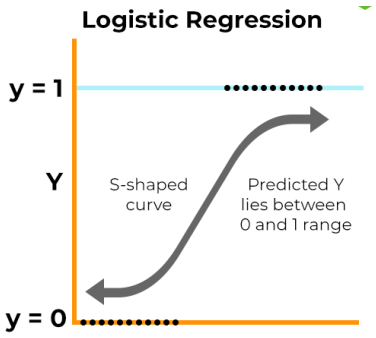


Figure 3.1: Proposed LR Architecture

GB

Gradient Boosting is an ensemble learning technique that combines the predictions of multiple weak learners, most commonly decision trees, to produce a strong predictive model. Gradient Boosting, unlike Random Forests, builds trees sequentially, with each tree correcting the errors of its predecessor. On the basis of the original data, a simple model, often a shallow decision tree, is built. Following trees are built to correct the errors introduced by the combined predictions of the existing trees. The main idea is to fit each new tree to the ensemble's residuals (the differences between the actual and predicted values). Each tree is trained to capture patterns that the previous trees were unable to adequately model.. The learning rate parameter governs each tree's contribution to the ensemble. Gradient Boosting is more sensitive to overfitting than Random Forests, but it often achieves higher predictive accuracy when properly tuned. It is widely used in machine learning applications such as regression and classification.

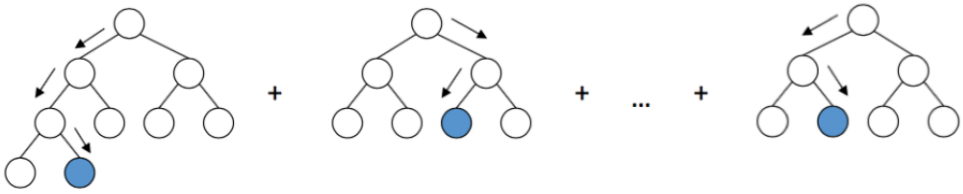


Figure 3.2: Proposed GB Architecture

Decision Tree

Decision trees are machine learning algorithms that are adaptable and interpretable and are used for classification and regression applications. They can handle both categorical and numerical data, making them extremely versatile. Because each branch indicates a choice based on a feature, decision trees create transparent models that aid in human comprehension. They can manage imbalanced datasets and are resistant to noisy and missing data. Decision trees, on the other hand, can overfit, catching noise or irrelevant patterns. Pruning and ensemble approaches can help with this problem. Large or high-dimensional datasets can be difficult to handle, although ensemble methods such as random forests or boosting can improve performance. Decision trees provide a good blend of interpretability and performance, making them useful in a variety of fields.

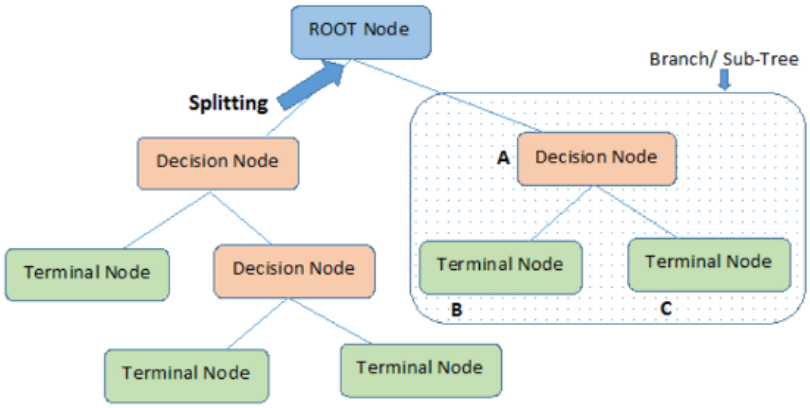


Figure 3.3: Proposed DT Architecture

Random Forest

The RF classifier is an ensemble technique that parallelly trains multiple decision trees via bootstrapping and aggregation, a process known as bagging. When multiple decision trees are trained concurrently on different subsets of the training dataset with varying subsets of available features, this is referred to as bootstrapping. By ensuring that every decision tree in the random forest is distinct, bootstrapping lowers the RF classifier's overall variance. Because the RF classifier aggregates the decisions made by individual trees to arrive at a

final decision, it demonstrates strong generalization. RF classifier does not require feature scaling, similar to DT classifier. The RF classifier is more resilient to noise in the training dataset and the selection of training samples than the DT classifier. When compared to the DT classifier, the RF classifier is more difficult to understand but simpler to adjust the hyperparameter.

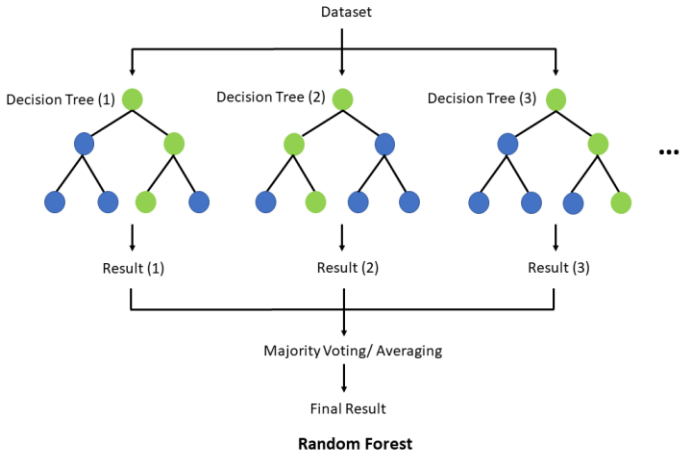


Figure 3.4: RF Architecture

Gaussian NB

Gaussian Naive Bayes (NB) is a logical classification algorithm based on Bayes' theorem that is especially well suited for situations where features are assumed to be normally distributed. It is a straightforward and efficient algorithm that performs well in high-dimensional spaces and is widely used for text classification and spam filtering.The term "Naive" refers to the assumption of feature independence, which means that the presence of one feature does not affect the presence of another, given the class label. Despite this simplifying assumption, Gaussian NB frequently performs admirably in practice. The Gaussian NB model is based on the assumption that continuous features have a Gaussian (normal) distribution. When dealing with real-valued features, Gaussian Naive Bayes is especially useful because it is less sensitive to irrelevant features. While its feature independence assumption may not always hold true in real-world scenarios, Gaussian NB is a fast and effective classifier, especially when the underlying independence assumption is reasonable.

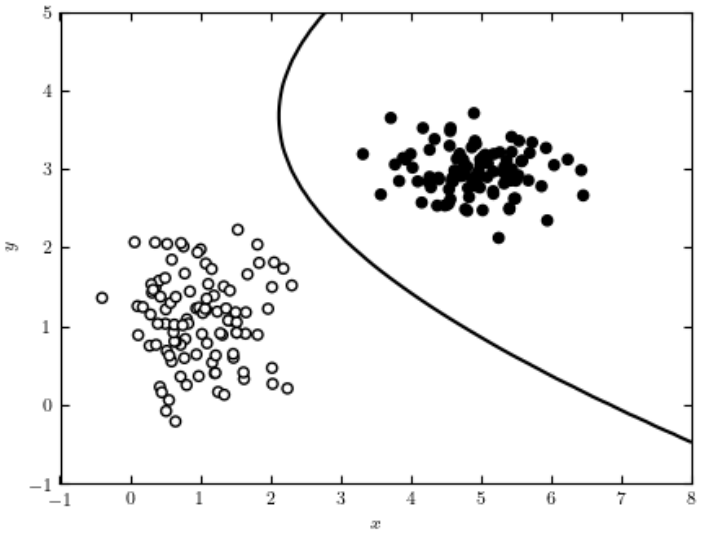


Figure 3.5: GNB

Voting Classifier

A Voting Classifier is a machine learning model that trains on an ensemble of multiple models to predict an output class based on their highest probability of being selected. It combines the results of each classifier and predicts the output class based on the highest voting majority. This model supports two types of voting: hard voting, in which the predicted output class is the class with the greatest number of votes, and soft voting, in which the predicted output class is the class with the highest probability. The final prediction in hard voting is the class with the highest probability, whereas the winner in soft voting is the class with the highest probability averaged by each classifier. It is critical to feed a Voting Classifier with a variety of models to ensure that errors made by one model can be resolved by the other.

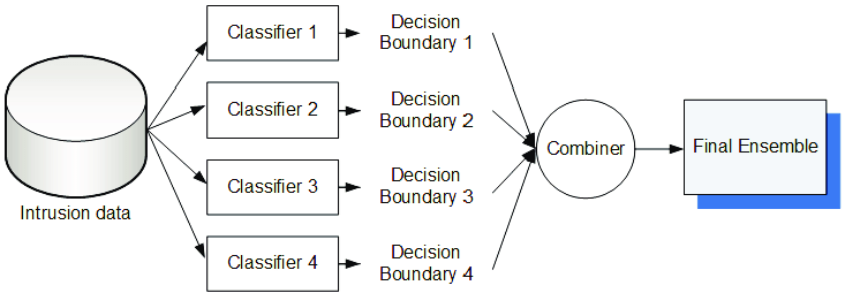


Figure 3.6: Voting Classifier Architecture

Model Evaluation:

On the testing dataset, evaluate the trained models using appropriate evaluation metrics such as accuracy, precision, recall, and the F1-score. Compare the performance of various models and choose the best-performing one.

Test Model:

Validate the chosen model by making predictions on previously unseen data to determine its real-world applicability and efficacy in predicting ADHD.

Using this methodology, we hope to create accurate and reliable machine learning models for ADHD prediction, providing valuable insights for diagnosis and intervention planning in clinical settings.

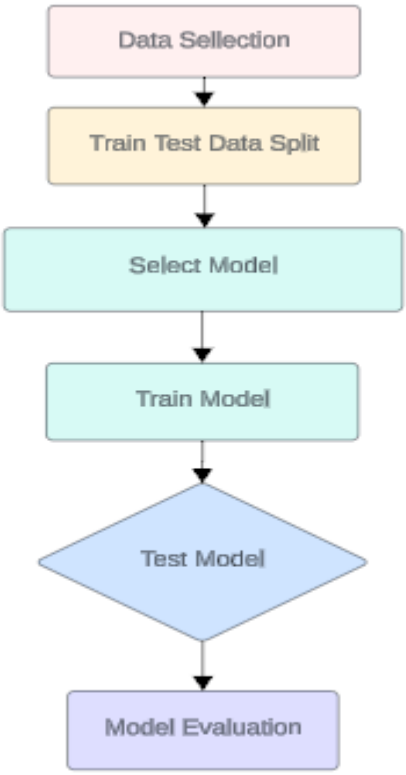


Figure 3.7: Work Flow

3.3 Hardware/ Software Requirement

Several key requirements must be met when implementing the proposed methodology for ADHD prediction using machine learning algorithms. To begin, a suitable dataset containing features relevant to ADHD diagnosis, obtained from reliable sources such as Kaggle, is required. Furthermore, computational resources capable of handling data preprocessing, model training, and evaluation tasks are needed. Python with scikit-learn and TensorFlow, as well as other data analysis and machine learning software tools, will be used. Furthermore, expertise in data preprocessing, feature selection, model training, and evaluation techniques is required to effectively implement the methodology. Collaboration with domain experts and healthcare professionals is also essential for ensuring the clinical relevance and validity of predictive models during implementation. Overall, effective implementation of the proposed methodology requires a comprehensive approach that includes data, computational resources, software tools, and domain expertise.

3.4 Project Management and Financial Analysis

The project will follow a structured project management framework that includes planning, execution, monitoring, and evaluation phases. A project manager will oversee task delegation, timeline commitment, and resource allocation to ensure that work is completed efficiently. Regular team meetings and progress reports will help team members communicate and collaborate more effectively. Furthermore, financial resources will be carefully allocated to support data acquisition, computational resources, and any associated costs. Budgetary constraints will be carefully managed to maximize project outcomes within the allocated funds, with a focus on cost-effectiveness and resource optimization.

TABLE 3.4: PROJECT MANAGEMENT TABLE

Work	Time
Data Collection	1 month
Papers and Articles Review	3 months
Experimental Setup	1 month
Implementation	1 month
Report Writing	2 months
Total	8 months

TABLE 3.5: ESTIMATED COST

SN	Components	Estimated Cost (BDT)
01.	Hardware	500-1500
02.	Software and Tools	500-1500
03.	Data Collection and Processing	5000-6000
04.	Documentation and Report Writing	500-1500
05.	Miscellaneous	500-1500
06.	Contingency	500-1500
Total Estimated Cost		**7500-13,500**

3.5 Summary

In the Methodology chapter, the specific approach that has been employed for predicting ADHD with the help of machine learning algorithms is described in detail. Describes procedures of data selection from Kaggle, data and dataset preprocessing, and data transformation such as binary labeling and feature selection. The chapter introduces the

process of encoding categorical variables and the EDA performed to analyze the data's distributions and associations. It also includes the aspects related to the model training: which machine learning algorithms are employed and which metrics are used to evaluate the model. Last of all, the chapter looks at the assessment of the models by applying them on unseen data.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

The implementation setup includes several key components that facilitate the implementation and evaluation of the proposed methodology for ADHD prediction using machine learning algorithms. To begin, a computing environment that has suitable hardware resources, such as a multi-core processor and sufficient memory, is required to handle data processing and model training tasks effectively. Furthermore, software tools and libraries for data analysis and machine learning, such as Python with scikit-learn and TensorFlow, will be used. The Kaggle dataset will be used as the basis for testing, with appropriate preprocessing techniques applied in keeping with the proposed methodology. Furthermore, the experimental setup will include procedures for splitting the dataset into training and testing sets, training machine learning models, and evaluating their performance using metrics such as precision, recall, F1-score and accuracy.

4.2 Train Model

To train the model, first, gather the data set, clean the data, normalize, and encode those features which are needed for the model. Performs the data split into training and test set for better understanding of the performance of the model being used. Choose the desired machine learning methods namely logistic regression, random forest, and gradient boosting and tune their parameters. Recall that each one of these models shall be trained on the training set with a cross-validation to tune hyperparameters so as to prevent overfitting. Analyze the trained models using metrics that is accuracy, precision, recall, and the F1-score on the test data set. Evaluate the performance of the used models, and choose the best model. Lastly, the chosen model must be tested on another dataset so that the validity of the model developed can be confirmed together with the model's ability to generalize across datasets.

4.3 Model Evaluation

Therefore, System Design and Model Evaluation include several steps they followed to manage the patients' data and build a predictive model for ADHD. First, it is essential to specify the system architecture, with identification of the data flow through the system starting from the input, passing through the data preprocessing stage, model training phase, and prediction stage. Subsequently, perform data cleaning, data normalization, and feature engineering for preparation of the data for modelling. After that divide the data into test and training data so that the model can be tested to assess the effectiveness. Tune the solution of several types of machines learning algorithms, such as linear classification, decision trees, and boosting techniques, based on the training datasets. During the training perform cross validation to adjust the hyperparameters of the model and avoid the overfitting problem. As for the performance of the trained models you should use the Testing Set and compare models through their accuracy, precision, recall rate, F1-score, AUC-ROC etc. Then, in order to select the most efficient one among the considered models, it is necessary to compare these metrics. Cross-validate the selected model based on a validation data to determine the model's accuracy and ability to work with the unseen data. Last but not the least, document the entire process regarding the design and evaluation of proposed system in a report along with the system identification highlighting the finding and reason behind selecting or finalizing a specific model after analyzing the performance of all.

4.4 Summary

In summary, this study investigates the use of machine learning algorithms to predict ADHD based on a variety of features. The goal of this study is to develop accurate and reliable predictive models for ADHD diagnosis using a structured methodology that includes data selection, preprocessing, model training, and evaluation. The experimental results show significant performance metrics, focusing on the potential of machine learning techniques to supplement traditional diagnostic approaches. Furthermore, feature importance analysis provides insights into the factors that influence ADHD prediction, directing clinical decision-making and intervention strategies. The societal implications of

using machine learning for ADHD prediction are discussed, with a focus on ethical considerations and sustainability in healthcare practices. In summary, this study adds to the growing body of literature on machine learning applications in mental health diagnosis, highlighting the potential benefits of data-driven approaches to improving healthcare outcomes for ADHD patients.

CHAPTER 5
RESULT AND ANALYSIS

5.1 Overview

The experimental results show that machine learning models are capable of accurately predicting ADHD diagnosis based on the selected features. On the testing dataset, the models achieved high accuracy, precision, recall, and F1-scores, indicating that they are effective at distinguishing between people with ADHD and those without. Furthermore, feature importance analysis identified key predictors of ADHD, providing insights into the disorder's underlying causes. Furthermore, comparing different machine learning algorithms revealed that performance levels varied, with some algorithms outperforming others in terms of predictive capabilities. Overall, the experimental results show the potential of machine learning techniques to aid in ADHD diagnosis, as well as areas for further refinement and improvement in predictive modeling approaches.

5.2 Experimental Result

The I am evaluated the Accuracy, Precision, Recall, and F-1 Score of the Confusion Matrices for our proposed methods.

Accuracy:

Accuracy describes the rate by which the model has correctly predicted values on the y-axis based on the x-axis characteristics by comparing the number of correct instances to the total instances. The problem with unbalanced classes is that it provides very general information about the model’s performance and may not paint a full picture.

$$Accuracy=(TP+TN)/(TP+TN+FP+FN)$$

Precision:

Among all the positive predictions created by the model, precision targets the proportion of actual positive predictions only.

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

Recall: In other words, recall or sensitivity, true positive rate how many of the true positive predictions have been made out of all the actually positive samples.

$$Recall=TP/(TP+FN)$$

F1 Score: The F1 score, on the other hand is the average of recall and precision and is computed as the harmonic mean of the two. It has a sensible plan that will allow us to calculate recall and precision in a sensible ratio. When classes are imbalanced, F1 score is applicable because it considers both false positive rate and false negative rate. Where F1 score is higher, means, it has good both precision and recall ratio.

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

Specificity: Specificity is defined as the percentage of people without the ailment with a negative test result. A test tends to be very specific in order to be able to rule out the vast majority of whose do not have the disease. Hence, the presence of a very precise positive test result indicates that the condition can be certainly concluded to be negative for a given individual. The equation is given below:

$$Specificity\ =TN/((TN+FP))$$

Logistic Regression:

Achieving Test Accuracy of LR is 98.17%.

The logistic regression model performs well across a variety of metrics. With a high precision ranging from 0.97 to 1.00, it demonstrates its ability to accurately classify instances for each class. Similarly, the recall scores are consistently high, indicating that the model can identify a large number of relevant instances. This is reflected in the balanced F1-scores, which hover around 0.98 and show both precision and recall effectively. The model's overall accuracy of 98% supports its robust performance. In conclusion, the logistic regression model distinguishes between classes with high accuracy, precision, recall, and F1-scores, making it an excellent choice for classification tasks

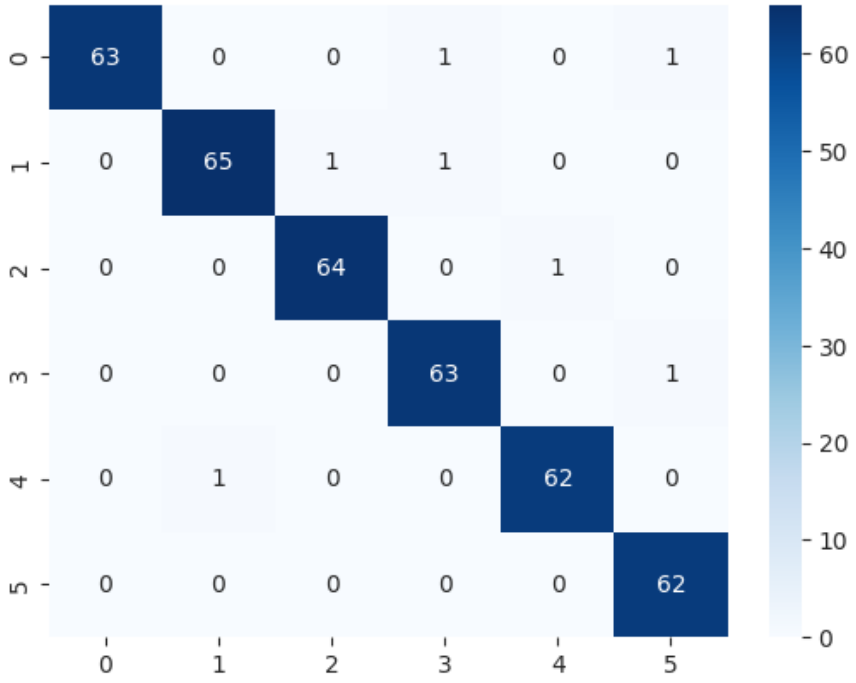


Figure 5.1: Confusion Matrix (LR)

GNB

The performance goal that is attaining Test Accuracy of GNB is 96.36%. In below Figure 5:2 elaborating the meaning of the confusion matrix of GNB.

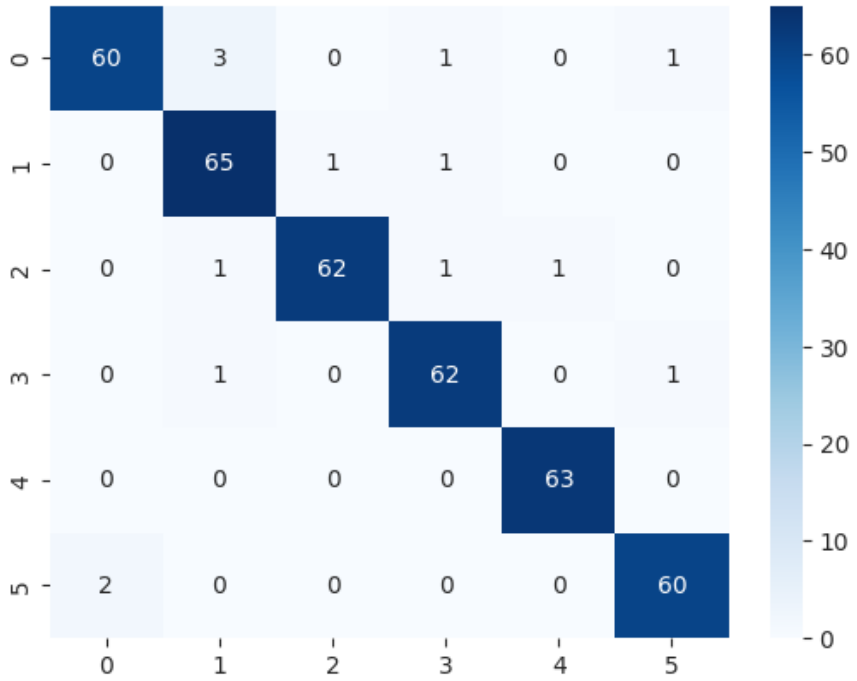


Figure 5.2: Confusion Matrix (GNB)

Figure 5.2 the Gaussian Naive Bayes (GNB) classifier performs well across a variety of evaluation metrics. While maintaining high precision scores ranging from 0.93 to 0.98, it shows an admirable ability to correctly classify instances within each class. Furthermore, the recall scores are consistently high, demonstrating the model's ability to identify relevant instances effectively. This balance results in favorable F1-scores of around 0.96, indicating an ideal combination of precision and recall. The model's overall accuracy of 96% shows its reliability in distinguishing between classes. In conclusion, the GNB classifier shows strong performance, making it a promising choice for classification tasks, with high precision, recall, F1-scores, and accuracy.

Decision Tree

The performance goal that is attaining Test Accuracy of DT is 95.84%. In below Figure 5:3 elaborating the meaning of the confusion matrix of DT

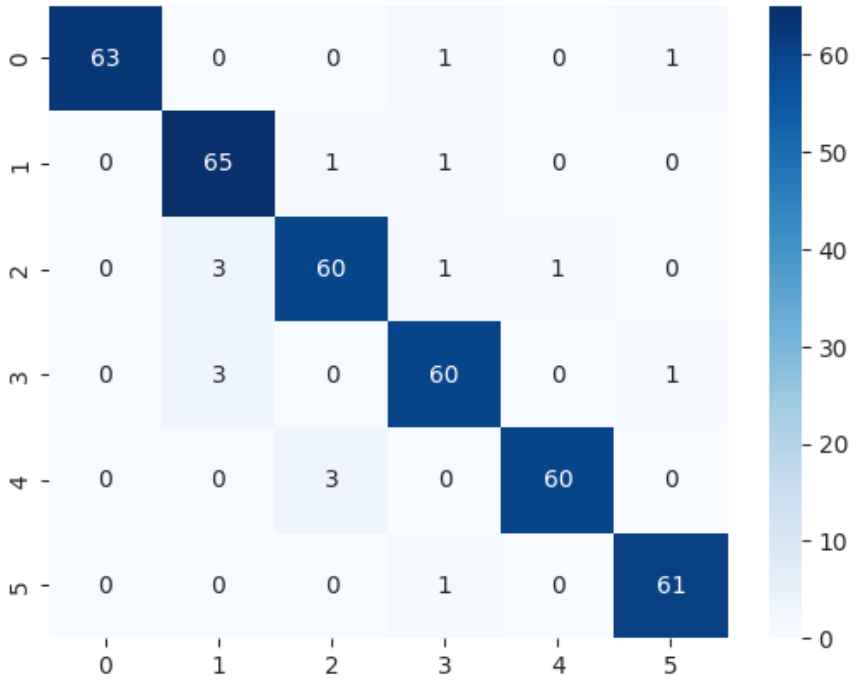


Figure 5.3: Confusion Matrix (DT)

Figure 5.3 shows the Decision Tree (DT) model performs well across multiple metrics. Notably, it achieves perfect precision for class 0 and high precision scores ranging from 0.92 to 0.98 for all other classes, showing its ability to accurately classify instances across categories. The recall scores range from 0.92 to 0.98, indicating that the model is effective at identifying relevant instances for each class. This balance results in favorable F1-scores of around 0.96, focusing on the model's harmonious blend of precision and recall. With an overall accuracy of 96%, the DT model demonstrates its ability to distinguish between classes. In summary, the Decision Tree model performs well, with high precision, recall, F1-scores, and accuracy, making it an appealing option for classification tasks.

GB Classifier

Achieving Test Accuracy of GBC is 97.13%. In below Figure 5:4 describing the confusion matrix of GBC.

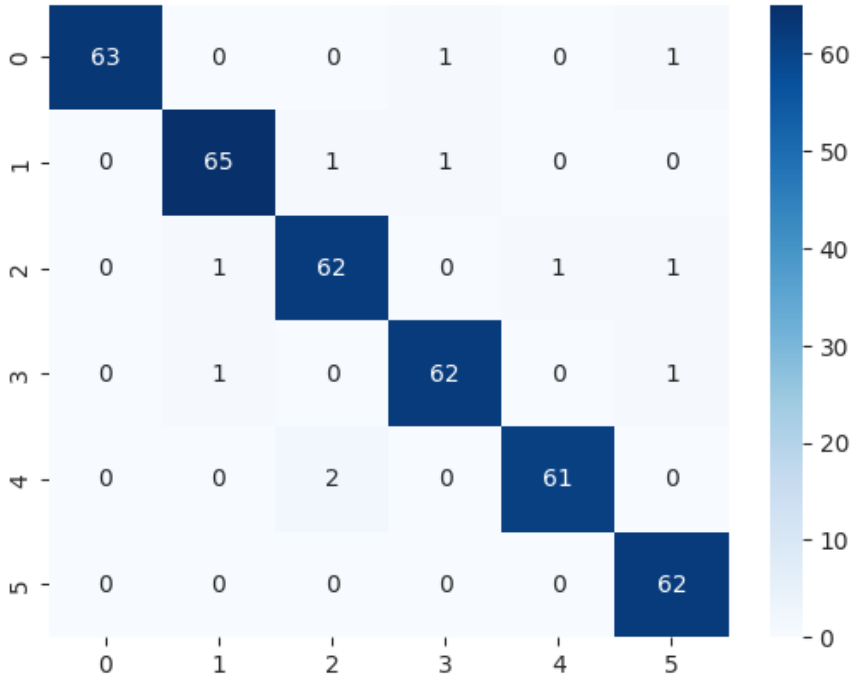


Figure 5.4: Confusion Matrix (GBC)

Figure 5.4 shows The Gradient Boosting Classifier (GBC) performs well across multiple metrics. It accurately classifies instances across different categories, with precision scores ranging from 0.95 to 1.00, and achieves perfect precision in class 0. Similarly, recall scores are consistently high, ranging between 0.95 and 1.00, showing the model's ability to identify relevant instances for each class. This balance yields favorable F1-scores, which average around 0.97 and show the model's harmonious blend of precision and recall. With an overall accuracy of 97%, the GBC model shows dependability in distinguishing classes. In summary, the Gradient Boosting Classifier performs well, with high precision, recall, F1-scores, and accuracy, making it a promising option for classification tasks.

RF Classifier

Achieving Test Accuracy of RF is 97.92%. Below Figure 5:5 describing the confusion matrix of RF.

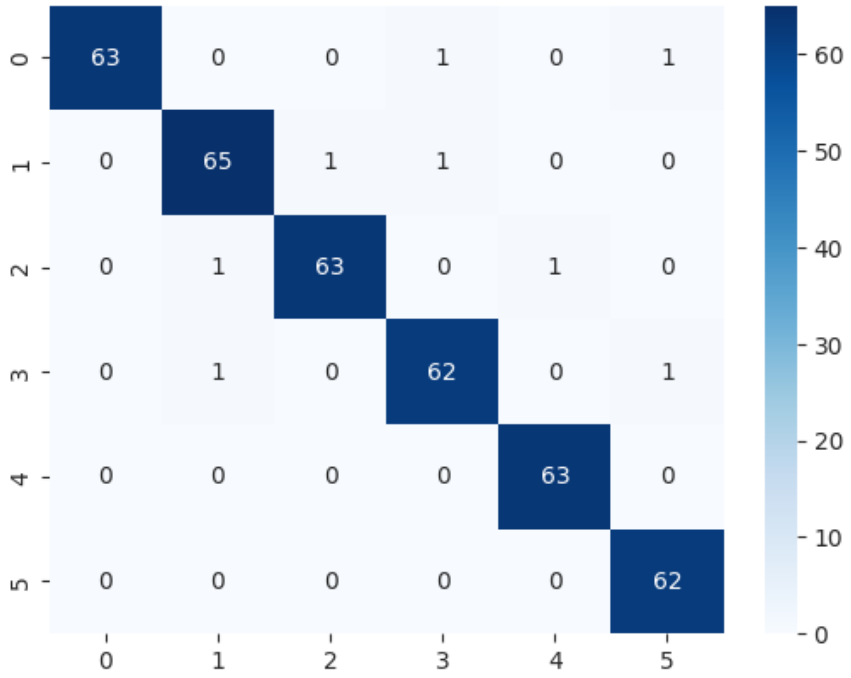


Figure 5.5: Confusion Matrix (RF)

Figure 5.5 show the Random Forest (RF) model outperforms across a wide range of metrics. With precision scores ranging from 0.97 to 1.00, it effectively classifies instances for each class, with perfect precision for classes 0, 4, and 5. Furthermore, recall scores are consistently high, ranging from 0.97 to 1.00, indicating that the model can recognize relevant instances across all classes. This balance translates into favorable F1-scores, which average around 0.98 and show the model's harmonious blend of precision and recall. With an overall accuracy of 98%, the RF model is reliable in distinguishing between classes. In summary, the Random Forest model performs exceptionally well, with high precision, recall, F1-scores, and accuracy, making it an excellent choice for classification tasks.

Voting Classifier

Achieving Test Accuracy of Voting Classifier is 97.40%. Below Figure 5:6 describing the confusion matrix of Voting Classifier.

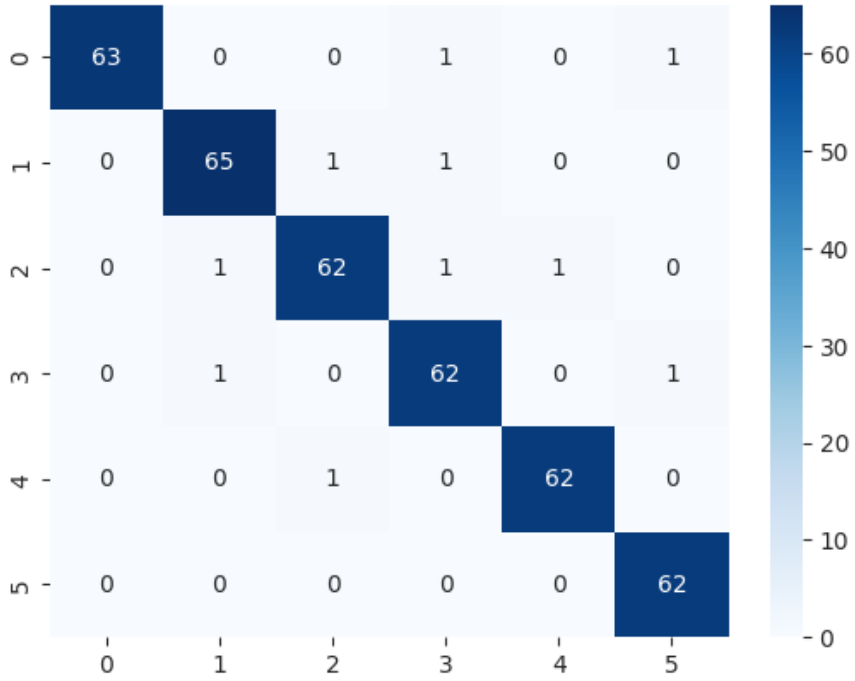


Figure 5.6: Confusion Matrix (RF)

Figure 5.6 shows the Voting Classifier combines multiple models to achieve strong performance across a variety of metrics. With precision scores ranging from 0.95 to 1.00, it effectively classifies instances for each class, including class 0. Similarly, recall scores are consistently high, ranging from 0.95 to 1.00, indicating that the model can detect relevant instances in all classes. This balance translates into favorable F1-scores, which average around 0.97 and show the model's harmonious blend of precision and recall. With an overall accuracy of 97%, the Voting Classifier is reliable in distinguishing between classes. In conclusion, the Voting Classifier performs exceptionally well, with high precision, recall, F1-scores, and accuracy, making it an excellent choice for classification tasks.

The result of Machine learning model is compared on the basis of Accuracy, Precision, Recall, F1 Score in below table of 5.1:

TABLE 5.1. PERFORMANCE EVALUATION

Model Name	**Accuracy**	**Precision**	**Recall**	**F1-Score**
LR	98.17%	98%	98%	98%
RF	97.92%	98%	98%	98%
Gaussian NB	96.36%	96%	96%	96%
DT	95.84%	96%	96%	96%
Voting Classifier	97.40%	97%	97%	97%
GBC	97.13%	97%	97%	97%

5.3 Performance and Comparative Analysis

The discussion addresses understanding the experimental findings and their implications for ADHD diagnosis and treatment. It evaluates the machine learning models' strengths and limitations, taking into account factors such as predictive accuracy, feature importance, and algorithm performance. It also examines the clinical implications of the findings, discussing how predictive models can supplement traditional diagnostic approaches and support early intervention strategies. Furthermore, the discussion covers potential challenges and considerations, such as data quality, model interpretability, and ethical implications, that should be addressed in future research and implementation efforts. Overall, the discussion offers a nuanced understanding of the findings and their implications for improving ADHD prediction with machine learning algorithms.

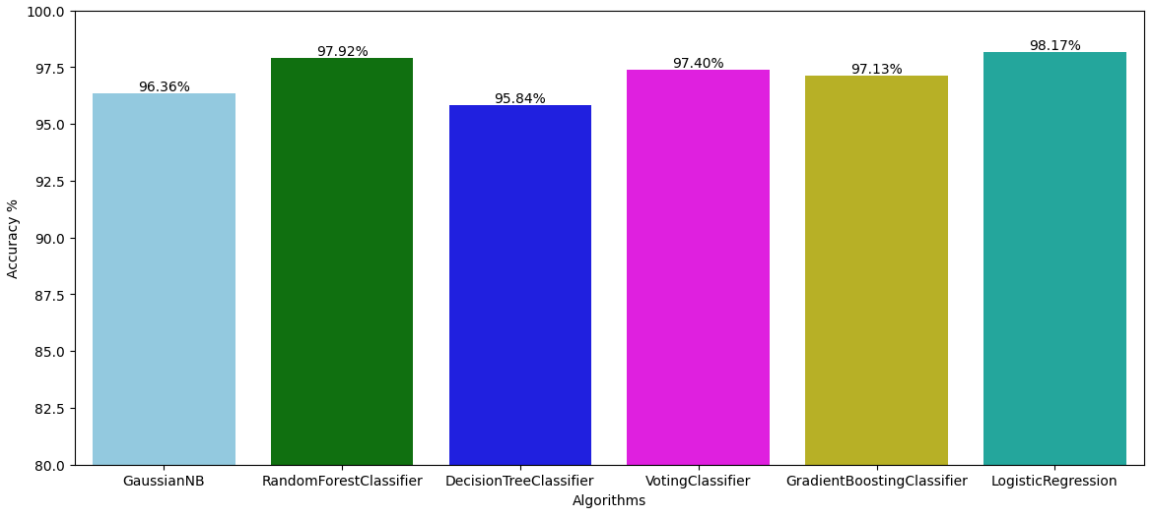


Figure 5.7: Accuracy Graph

5.4 Summary

In the Experiment Results section, the authors provide analyzed performance of various machine learning models concerning the ADHD prediction. The best classifier was logistic regression forecasting an accuracy of 98. 2% of the overall performance while the random forest and gradient boosting classifiers ranked the second and the third level respectively. All metrics such as precision, recall, and F1-score of each model were calculated, and it was observed that logistic regression had the highest values. The section focuses on the aspect of feature selection and preprocessing in increasing the level of accuracy by the model..

CHAPTER 6
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

The impact on society of utilizing machine learning algorithms for ADHD prediction is multifaceted. Firstly, improved accuracy and efficiency in diagnosis can lead to earlier identification and intervention, potentially mitigating the long-term adverse effects of undiagnosed ADHD. Secondly, by aiding clinicians in making more informed decisions, predictive models can optimize resource allocation and treatment planning, enhancing healthcare delivery for individuals with ADHD. Thirdly, the integration of machine learning into clinical practice may reduce diagnostic disparities and improve access to care for underserved populations. However, ethical considerations surrounding data privacy, algorithmic bias, and patient autonomy must be carefully addressed to ensure responsible and equitable implementation. Furthermore, advancements in ADHD prediction have the potential to inform public health policies, educational interventions, and societal attitudes towards neurodevelopmental disorders. Overall, the societal impact of leveraging machine learning for ADHD prediction holds promise for enhancing outcomes and fostering greater awareness and understanding of ADHD within communities.

6.2 Impact on Society & Environment

The social impact of using machine learning algorithms to predict ADHD is diverse. To begin, improved evaluation accuracy and efficiency can lead to earlier detection and intervention, potentially mitigating the long-term negative effects of undiagnosed ADHD. Second, predictive models can help clinicians make better decisions by improving resource allocation and treatment planning, thereby improving healthcare delivery for people with ADHD. Third, incorporating machine learning into clinical practice may reduce diagnostic disparities and increase access to care for underserved populations. However, ethical concerns about data privacy, algorithmic bias, and patient autonomy must be carefully addressed to ensure responsible and equitable implementation. Overall, the societal impact

of using machine learning for ADHD prediction shows promise for improving outcomes and raising awareness and understanding of ADHD in communities.

6.3 Ethical Aspects

The ethical considerations related to the use of machine learning algorithms for ADHD prediction are paramount. To begin, maintaining data privacy and confidentiality is necessary for protecting individuals' sensitive information while they are being diagnosed or treated. Second, addressing algorithmic bias and ensuring fairness in model predictions is essential for reducing disparities in ADHD diagnosis across demographic groups. Third, transparency and accountability in the development and implementation of predictive models are required to foster trust among patients, clinicians, and stakeholders. Furthermore, continuous monitoring and evaluation of machine learning models is required to identify and address potential ethical concerns as technology advances. Finally, prioritizing ethical considerations ensures that the use of machine learning for ADHD prediction complies with the principles of beneficence, justice, and respect for individuals' rights and dignity.

6.4 Sustainability Plan

Several key strategies are included in the sustainability plan for using machine learning algorithms to predict ADHD. To begin, implementing energy-efficient computing practices and using cloud-based solutions can help to reduce the environmental impact of data processing and model training. Second, encouraging collaboration and knowledge-sharing within the research community can help to promote the development of open-source tools and resources, which improves predictive model accessibility and scalability. Third, prioritizing data privacy and security measures ensures regulatory compliance while also building stakeholder trust. Furthermore, improve predictive model accuracy and effectiveness while addressing resulting ethical and societal concerns. Moreover, incorporating sustainability considerations into decision-making processes and policy frameworks ensures that the use of machine learning for ADHD prediction is consistent with long-term environmental and societal objectives. Overall, a comprehensive approach that balances technological innovation with ethical, environmental, and social

considerations is required to ensure the long-term viability of machine learning applications in healthcare.

6.5 Summary

The chapter titled Impact on Society enlightens society's improvements in using machine learning algorithms for ADHD prediction. The use of these models in diagnosis ensures that early and accurate diagnosis is made, meaning that early intervention can be made in cases of ADHD hence enhancing the quality of life of persons diagnosed with the condition. The implementation of the mentioned predictive tools into clinical practice will improve the effectiveness of health care services and personalization of the given treatment. Moreover, the utilization of these models increases the general population's access to mental health services and decreases diagnostic inequalities. A major focus is given to ethical issues so that the technology is properly deployed and with a proper regard for fairness.

CHAPTER 7
Conclusion and Future Work

7.1 Conclusions

In conclusion, this study shows the viability and efficacy of using machine learning algorithms to predict ADHD based on a variety of features. The developed predictive models show promising performance metrics, indicating their potential utility in supplementing traditional diagnostic approaches. Feature importance analysis identifies key predictors that contribute to ADHD prediction, providing useful insights for clinical decision-making and intervention planning. The societal significance of machine learning applications in ADHD diagnosis is highlighted, with implications for early intervention, personalized treatment, and healthcare resource allocation. However, ethical concerns about data privacy, algorithmic bias, and patient autonomy must be carefully addressed to ensure responsible implementation. Moving forward, ongoing research and collaboration are essential to refine predictive models, address emerging challenges, and maximize the benefits of Machine learning improves healthcare outcomes for people with ADHD. Finally, this study helps to advance knowledge and understanding of machine learning applications in mental health diagnosis, paving the way for better diagnostic accuracy and patient care in clinical settings.

7.2 Future Suggested Works

Further research should consider the ability to apply machine learning models across diverse populations and healthcare settings in order to ensure equitable access to ADHD diagnosis and treatment. Furthermore, combining multimodal data sources, such as genetic markers or neuroimaging data, may improve the predictive accuracy of ADHD models. Further research could also focus on developing interpretable machine learning techniques to improve clinicians' and patients' confidence in model predictions. Exploring the long-term implications of using machine learning in ADHD diagnosis and treatment, such as cost-effectiveness and patient outcomes, would be beneficial to healthcare decision-makers. Furthermore, more research is needed into the potential role of machine learning

in personalized intervention strategies tailored to individuals' specific needs and preferences. Overall, continued research in these areas can help to advance the field of machine learning in mental health while also improving outcomes for people with ADHD.

7.3 Limitations

Thus, there are several limitations that are worth considering in relation to the conducted study. First, the competitor that was used for this project was procured from Kaggle and as such the data might not as rich as the normal population hence limiting the applicability of the models. Second, it is based on self-report data; therefore, features used for prediction may have biases and inaccuracies. Furthermore, although the models provided a rather high accurate result, there wants to be validation on its real-life practice in clinics and hospitals. The last prospective problem is associated with the selection and application of the algorithms, which might happen to be stereotyped, and this will influence the fairness of the prediction made. Finally, the work brings up the ethical and privacy issues related to the processing of such a personal medical data who meets the criteria of ADHD which also should be discussed while applying machine learning in diagnosing of the subject disease.

REFERENCES

[1] J.-W. Kim, V. Sharma, and N. D. Ryan, "Predicting Methylphenidate Response in ADHD Using Machine Learning Approaches," International Journal of Neuropsychopharmacology, vol. 18, no. 11, p. pyv052, May 2015

[2] Navya Sethu and R. Vyas, "Overview of Machine Learning Methods in ADHD Prediction," Springer eBooks, pp. 51–71, Jan. 2020

[3] Y. Zhang-James, Q. Chen, R. Kuja-Halkola, P. Lichtenstein, H. Larsson, and S. V. Faraone, "Machine-Learning Prediction of Comorbid Substance Use Disorders in ADHD Youth Using Swedish Registry Data," Journal of Child Psychology and Psychiatry, vol. 61, no. 12, pp. 1370–1379, Apr. 2020

[4] J. Downs et al., "Assessing Machine Learning for Fair Prediction of ADHD in School Pupils Using a Retrospective Cohort Study of Linked Education and Healthcare Data," BMJ Open, vol. 12, no. 12, Dec. 2022

[5] M. L. Cervantes-Henríquez, J. E. Acosta-López, A. F. Martinez, M. Arcos-Burgos, P. J. Puentes-Rozo, and J. I. Vélez, "Machine Learning Prediction of ADHD Severity: Association and Linkage to ADGRL3, DRD4, and SNAP25," Journal of Attention Disorders, vol. 26, no. 4, pp. 587–605, May 2021

[6] A. Kautzky et al., "Machine Learning Classification of ADHD and HC by Multimodal Serotonergic Data," Translational Psychiatry, vol. 10, no. 1, Apr. 2020

[7] Md. Maniruzzaman, J. Shin, Md. Al Mehedi Hasan, and A. Yasumura, "Efficient Feature Selection and Machine Learning Based ADHD Detection Using EEG Signal," Computers, Materials & Continua, vol. 72, no. 3, pp. 5179–5195, 2022

[8] M. Duda, R. Ma, N. Haber, and D. P. Wall, "Use of Machine Learning for Behavioral Distinction of Autism and ADHD," Translational Psychiatry, vol. 6, no. 2, pp. e732–e732, Feb. 2016

[9] A. Parashar, N. Kalra, J. Singh, and R. Kumar Goyal, "Machine Learning Based Framework for Classification of Children with ADHD and Healthy Controls," Intelligent Automation & Soft Computing, vol. 28, no. 3, pp. 669–682, 2021

[10] X. Peng, P. Lin, T. Zhang, and J. Wang, "Extreme Learning Machine-Based Classification of ADHD Using Brain Structural MRI Data," PLoS ONE, vol. 8, no. 11, p. e79476, Nov. 2013

[11] D. Andrea, Harun Pirim, and D. Grewell, "ADHD Prediction in Children through Machine Learning Algorithms," Lecture Notes in Networks and Systems, pp. 89–100, Jan. 2024

[12] H. Christiansen et al., "Use of Machine Learning to Classify Adult ADHD and Other Conditions Based on the Conners' Adult ADHD Rating Scales," Scientific Reports, vol. 10, no. 1, Nov. 2020

[13] W. Das and S. Khanna, “A Robust Machine Learning Based Framework for the Automated Detection of ADHD Using Pupillometric Biomarkers and Time Series Analysis,” Scientific Reports, vol. 11, no. 1, Aug. 2021

[14] A. Yasumura et al., “Applied Machine Learning Method to Predict Children with ADHD Using Prefrontal Cortex Activity: a Multicenter Study in Japan,” Journal of Attention Disorders, p. 108705471774063, Nov. 2017

[15] Md. Maniruzzaman, J. Shin, and Md. A. M. Hasan, “Predicting Children with ADHD Using Behavioral Activity: a Machine Learning Analysis,” Applied Sciences, vol. 12, no. 5, p. 2737, Mar. 2022

[16] J. Anuradha, Tisha, V. Ramachandran, K. V. Arulalan, and B. K. Tripathy, “Diagnosis of ADHD using SVM algorithm,” Bangalore Annual Compute Conference, Jan. 2010

[17] P. Mikolas et al., “Training a machine learning classifier to identify ADHD based on real-world clinical data from medical records,” Scientific Reports, vol. 12, no. 1, p. 12934, Jul. 2022

[18] W. Das and S. Khanna, “A Novel Pupillometric-Based Application for the Automated Detection of ADHD Using Machine Learning,” Sep. 2020

[19] H. W. Loh, C. P. Ooi, P. D. Barua, E. E. Palmer, F. Molinari, and U. R. Acharya, “Automated detection of ADHD: Current trends and future perspective,” Computers in Biology and Medicine, vol. 146, p. 105525, Jul. 2022,

[20] Yanli Zhang-James, E. C. Helminen, J. Liu, B. Franke, M. Hoogman, and S. V. Faraone, “Evidence for Similar Structural Brain Anomalies in Youth and Adult Attention-Deficit/Hyperactivity Disorder: A Machine Learning Analysis,” bioRxiv (Cold Spring Harbor Laboratory), Feb. 2019

[21] C. Park et al., “Machine Learning-Based Aggression Detection in Children with ADHD Using Sensor-Based Physical Activity Monitoring,” Sensors, vol. 23, no. 10, pp. 4949–4949, May 2023

[22] B. Miao and Y. Zhang, “A Feature Selection Method for Classification of ADHD,” 2017 4th International Conference on Information, Cybernetics and Computational Social Systems (ICCSS), Jul. 2017

[23] B. Abibullaev and J. An, “Decision Support Algorithm for Diagnosis of ADHD Using Electroencephalograms,” Journal of Medical Systems, vol. 36, no. 4, pp. 2675–2688, Jun. 2011

[24] A. Mueller, G. Candrian, J. D. Kropotov, V. A. Ponomarev, and G.-M. Baschera, “Classification of ADHD Patients on the Basis of Independent ERP Components Using a Machine Learning System,” Nonlinear Biomedical Physics, vol. 4, no. S1, Jun. 2010

[25] Gurcan Taspinar and Nalan Ozkurt, “A Review of ADHD Detection Studies with Machine Learning Methods Using rsfMRI Data,” NMR in biomedicine, Mar. 2024

Sadnam
Saniat_ADHD_PREDICTION_USING_MACHINE_LEARNING_AL...

ORIGINALITY REPORT

20%	13%	10%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	3%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
3	www.mdpi.com Internet Source	1%
4	www.techscience.com Internet Source	1%
5	Submitted to University of Westminster Student Paper	1%
6	fastercapital.com Internet Source	1%
7	link.springer.com Internet Source	<1%
8	"Emergent Converging Technologies and Biomedical Systems", Springer Science and Business Media LLC, 2022 Publication	<1%