DISEASE PREDICTION FOR POLYCYSTIC OVARY SYNDROME (PCOS) USING DEEP LEARNING AND CONVENTIONAL MACHINE LEARNING ALGORITHMS

BY

Soniya Akter
ID: 212-15-4228

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

Supervised By

Md. Sazzadur Ahamed
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Naznin Sultana
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

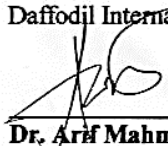
16 JULY, 2024

APPROVAL

This Project titled "**DISEASE PREDICTION FOR POLYCYSTIC OVARY SYNDROME (PCOS) USING DEEP LEARNING AND CONVENTIONAL MACHINE LEARNING ALGORITHMS**", submitted by **Soniya Akter** and to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 16 July, 2024.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque — **Board Chairman**
Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University



Dr. Arif Mahmud — **Internal Examiner 1**
Associate Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

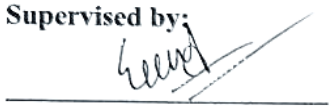
Md. Sazzadur Ahamed — **Internal Examiner 2**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

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

DECLARATION

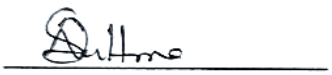
I hereby declare that, this project has been done by us under the supervision of **Md. Sazzadur Ahamed, Assistant Professor, and Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any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

Supervised by:



Md. Sazzadur Ahamed
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:



Naznin Sultana
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Soniya Akter
ID: 212-15-4228
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

We begin by extending our heartfelt gratitude to the Almighty for blessing us and enabling us to successfully complete our final year project/internship.

Our sincere appreciation goes to **Md. Sazzadur Ahamed, Assistant Professor** of the Department of Computer Science and Engineering at Daffodil International University, Dhaka. Their extensive knowledge and keen interest in the field of "Machine Learning" were instrumental in guiding us throughout this project. Their unwavering patience, scholarly guidance, continuous encouragement, dedicated supervision, constructive criticism, valuable advice, and thorough review of multiple drafts at every stage played a crucial role in the completion of this project.

We would also like to express our deepest thanks to **Dr. Sheak Rashed Haider Noori**, the Professor & Head of the Department of CSE, for his invaluable assistance in bringing our project to fruition. Our gratitude extends to all the faculty members and staff of the CSE department at Daffodil International University.

We are grateful to our fellow classmates at Daffodil International University, who engaged in discussions and provided support during the course of our project.

Lastly, we would like to acknowledge and show our utmost respect for the unwavering support and patience of our parents.

ABSTRACT

In recent years, there has been a noticeable rise in the prevalence of Polycystic Ovary Syndrome (PCOS), a complex endocrine disorder that affects a significant portion of the population, particularly women of reproductive age. PCOS is characterized by hormonal imbalances, irregular menstrual cycles, and the presence of multiple small cysts on the ovaries. Beyond its reproductive implications, PCOS is associated with various metabolic disturbances, including insulin resistance, obesity, dyslipidemia, and increased risk for type 2 diabetes and cardiovascular disease. To gain a comprehensive understanding of the disease's severity and its multifaceted impact on women's health, distinguishing between standard and affected diagnostic reports is imperative. In this study, we propose the application of algorithmic models to enable early detection and raise awareness of potential health risks associated with PCOS. Our approach is straightforward and well-suited for the prediction of uncomplicated cases of PCOS in real-world scenarios. Our dataset, sourced from various medical databases and clinical records, served as the foundation for our research. We employed a wide array of classifiers, including ANN, RNN, CNN, LSTM, BLSTM, RF, LR, GB, KNN, ABC, DT, SVM, QDA, RC, PA, GNB, and ensemble techniques, to comprehensively explore and evaluate the predictive capabilities of each model in identifying PCOS and its associated complications. The results yielded notable success, with the Soft Voting classifier emerging as the most accurate, an impressive accuracy rate of 96.58%. Our optimization efforts, which included hyperparameter tuning, further enhanced the performance of each classifier. Based on extensive experimentation and a review of contemporary research, our findings unequivocally endorse the Support Vector Classifier (SVC) classifier as exceptionally proficient, demonstrating a remarkable accuracy rate of 96.50% in the precise prediction of PCOS disease.

Keywords: PCOS, Bagging, Boosting, Ensemble, Machine Learning, Deep Learning.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Polycystic Ovary Syndrome (PCOS) stands as one of the most prevalent endocrine disorders affecting women worldwide. With its intricate interplay of hormonal imbalances, metabolic disturbances, and reproductive irregularities, PCOS presents a multifaceted clinical challenge. This syndrome, characterized by its complex pathophysiology and diverse clinical manifestations, poses significant diagnostic and management dilemmas for healthcare providers. As a syndrome with variable clinical presentation and evolving diagnostic criteria, PCOS has garnered increasing attention in recent decades, not only due to its high prevalence but also due to its profound impact on women's health and quality of life.

At its core, PCOS is a heterogeneous disorder, encompassing a spectrum of symptoms and metabolic abnormalities that extend beyond its reproductive manifestations. Historically recognized for its association with menstrual irregularities, anovulation, and infertility, PCOS now stands as a syndrome with systemic implications, affecting various organ systems and posing long-term health risks. The hallmark features of PCOS include hyperandrogenism, ovarian dysfunction, and polycystic ovarian morphology, although the presentation of these features can vary widely among affected individuals. Furthermore, the syndrome is frequently accompanied by metabolic disturbances, including insulin resistance, dyslipidemia, obesity, and an increased risk for type 2 diabetes mellitus and cardiovascular disease. These metabolic aberrations not only exacerbate the reproductive manifestations of PCOS but also contribute to its long-term morbidity and mortality, underscoring the importance of early detection and comprehensive management strategies.

Given the heterogeneous nature of PCOS and its varied clinical manifestations, accurate diagnosis and effective management pose significant challenges for healthcare providers. The diagnosis of PCOS relies on a combination of clinical, biochemical, and imaging

criteria, as outlined by various expert consensus guidelines. However, the lack of a singular diagnostic test and the evolving nature of diagnostic criteria contribute to diagnostic uncertainty and may lead to underdiagnosis or misdiagnosis of the syndrome. Moreover, the clinical heterogeneity of PCOS necessitates a personalized approach to management, tailored to the individual patient's presentation, reproductive goals, and metabolic profile. Thus, there is a growing need for reliable predictive models and decision-support tools that can facilitate early detection, risk stratification, and personalized management of PCOS.

1.2 Motivation

In recent years, the advent of machine learning and artificial intelligence (AI) has revolutionized the field of healthcare, offering innovative solutions for disease prediction, diagnosis, and prognostication. Leveraging vast datasets and sophisticated algorithms, machine learning models have demonstrated remarkable capabilities in identifying complex patterns and relationships within clinical data, thereby enabling accurate disease prediction and risk stratification. In the context of PCOS, machine learning holds immense promise as a tool for early detection, prognosis, and personalized management, offering the potential to enhance clinical decision-making and improve patient outcomes.

By harnessing the power of machine learning algorithms and integrating multidimensional clinical data, researchers aim to develop robust predictive models for PCOS that can effectively discriminate between affected and unaffected individuals, identify at-risk populations, and predict disease progression and associated complications. These models utilize a variety of input variables, including demographic data, clinical symptoms, hormonal profiles, imaging findings, and genetic markers, to generate predictive algorithms that are tailored to the heterogeneous nature of PCOS. Moreover, the extraction of significant insights from complicated datasets is made possible by machine learning approaches like feature selection, dimensionality reduction, and ensemble learning, which improve the predicted accuracy and interpretability of the models.

The motivation behind developing a disease prediction model for Polycystic Ovary Syndrome (PCOS) using both Deep Learning and Conventional Machine Learning

algorithms is the result of the urgent necessity to handle the difficulties in identifying and treating this complicated endocrine condition.

PCOS affects millions of women worldwide and is characterized by hormonal imbalances, ovarian cysts, irregular menstrual cycles, and potential complications such as infertility, diabetes, and cardiovascular diseases. Despite its prevalence and impact on women's health, PCOS often goes undiagnosed or misdiagnosed due to its varied clinical presentation and lack of specific diagnostic tests.

By leveraging the power of advanced computational techniques like Deep Learning and Conventional Machine Learning, we aim to develop a robust and accurate predictive model that can assist healthcare professionals in early detection and personalized management of PCOS. Such a model holds the potential to revolutionize clinical practice by:

- **Early Detection:** By analyzing diverse datasets including clinical symptoms, hormonal profiles, and imaging studies, the model can identify patterns indicative of PCOS at an early stage, allowing for timely intervention and preventive measures.
- **Personalized Medicine:** Tailoring treatment strategies to individual patients based on their unique risk factors, genetic predispositions, and response to therapy can significantly improve outcomes and reduce the burden of PCOS-related complications.
- **Improved Patient Care:** Providing healthcare providers with reliable decision support tools can enhance diagnostic accuracy, facilitate patient counseling, and streamline the referral process to specialists when necessary, leading to more effective and efficient healthcare delivery.
- **Research Advancements:** Insights gained from analyzing large-scale datasets using machine learning algorithms can deepen our understanding of PCOS pathophysiology, identify novel biomarkers, and inform the development of targeted therapies and preventive interventions.

By bridging the gap between cutting-edge technology and clinical practice, our initiative aims to empower both patients and healthcare providers in the fight against PCOS, ultimately improving the quality of life for millions of women worldwide.

1.3 Rationale of the Study

The development of accurate predictive models for PCOS holds immense potential to revolutionize clinical practice, offering clinicians valuable tools for early detection, risk assessment, and personalized management of this complex syndrome. By leveraging the vast amount of clinical data available and harnessing the power of machine learning algorithms, researchers can unlock novel insights into the pathophysiology of PCOS, identify novel biomarkers, and develop targeted interventions aimed at mitigating the long-term health risks associated with the syndrome. Moreover, by empowering patients with timely diagnosis and personalized management strategies, predictive models for PCOS can improve patient outcomes, enhance quality of life, and reduce the burden of this prevalent endocrine disorder on healthcare systems globally.

A large number of academic institutes have worked to create machine learning algorithms for PCOS detection. But it turned out that their techniques were insufficient and inaccurate in their predictions. Our plan is to improve the body's ability to predict illness. The two main categories of machine learning techniques are supervised and unsupervised learning. Supervised learning makes use of the dataset's training data and labeled data to produce outputs from inputs based on predetermined associations. On the other hand, unsupervised learning builds models by utilizing unlabeled data to reveal hidden patterns and information. Our method seeks to forecast a person's risk of developing PCOS, providing a useful instrument for early identification and preventive care.

Our research introduces a model for predicting PCOS, a condition on the rise and significantly affecting our society. In our resource-constrained setting, the scarcity of diagnostic tools and high costs of analysis pose challenges for identifying this ailment through traditional means. To address this issue, we turn to machine learning, leveraging

its capabilities to analyze symptoms and enhance diagnostic accuracy, thereby offering a cost-effective solution for our community.

The rationale behind conducting a study on disease prediction for Polycystic Ovary Syndrome (PCOS) using Deep Learning and Conventional Machine Learning algorithms lies in the need to address several critical aspects of this prevalent yet often misunderstood condition:

Diagnostic Challenges: PCOS diagnosis is notoriously complex due to its heterogeneous presentation and overlapping symptoms with other conditions. Conventional diagnostic criteria rely on a combination of clinical manifestations, biochemical markers, and imaging studies, making it prone to subjectivity and misinterpretation. Developing a reliable predictive model can assist in early detection and accurate diagnosis, leading to timely interventions and improved patient outcomes.

Personalized Medicine: PCOS is a multifaceted disorder with significant variability in clinical phenotypes and treatment responses among affected individuals. A tailored approach to patient management based on individual risk factors, genetic predispositions, and metabolic profiles is crucial for optimizing therapeutic outcomes and preventing long-term complications. By integrating machine learning algorithms into clinical practice, we can stratify patients into distinct subgroups and deliver personalized care plans tailored to their specific needs.

Data-driven Insights: The proliferation of electronic health records (EHRs), genomic data, and imaging studies has generated vast amounts of healthcare data ripe for analysis. Leveraging advanced computational techniques such as Deep Learning allows us to extract meaningful insights from these diverse datasets, uncover hidden patterns, and identify novel biomarkers associated with PCOS pathophysiology. By elucidating the underlying mechanisms driving PCOS development and progression, our study aims to inform the development of targeted interventions and preventive strategies.

Clinical Decision Support: Healthcare providers often face challenges in interpreting complex clinical data and making informed diagnostic and therapeutic decisions in real-

time. Integrating machine learning models into clinical workflows can provide decision support tools that augment clinical reasoning, improve diagnostic accuracy, and enhance treatment planning. By empowering healthcare providers with evidence-based recommendations, our study aims to facilitate more efficient and effective patient care delivery while minimizing the risk of diagnostic errors and treatment delays.

1.4 Research Question

1. How can machine learning aid in early detection of PCOS?
2. What are the key features for predicting PCOS risk?
3. How accurate are Deep Learning algorithms in diagnosing PCOS compared to traditional methods?
4. Can machine learning models predict long-term complications of PCOS?
5. How can the integration of clinical, biochemical, and imaging data improve PCOS prediction accuracy?

1.5 Expected output

The expected output of the study is a validated predictive model that leverages Deep Learning and Conventional Machine Learning algorithms to accurately detect Polycystic Ovary Syndrome (PCOS) and facilitate personalized management strategies. This model is anticipated to provide healthcare professionals with a reliable tool for early identification of PCOS risk, aiding in timely interventions and improving patient outcomes. Additionally, the study aims to enhance our understanding of PCOS pathophysiology, identify novel biomarkers, and optimize treatment plans tailored to individual patient characteristics. Ultimately, the development and validation of this predictive model hold the potential to revolutionize PCOS diagnosis and management, leading to more effective healthcare delivery and improved quality of life for individuals affected by this complex endocrine disorder.

1.6 Project Management and Finance

1.6.1 Project Management:

The project will follow a structured approach to ensure efficient execution and timely completion. It will involve the following key steps:

1.6.1.1 Planning: Define project objectives, scope, and deliverables. Develop a detailed project plan outlining tasks, milestones, and timelines.

1.6.1.2 Resource Allocation: Identify and allocate necessary resources including personnel, data sources, computing infrastructure, and software tools.

1.6.1.3 Team Formation: Assemble a multidisciplinary team comprising data scientists, medical professionals, and domain experts to collaborate on model development and validation.

1.6.1.4 Data Collection and Preprocessing: Gather relevant datasets including clinical records, hormonal profiles, imaging studies, and genetic data. Preprocess and clean the data to ensure quality and consistency.

1.6.1.5 Model Development: Design and implement Deep Learning and Conventional Machine Learning algorithms for PCOS prediction. Experiment with different model architectures, feature selection techniques, and hyperparameters optimization.

1.6.1.6 Validation and Evaluation: Evaluate the performance of the predictive model using independent datasets through rigorous validation techniques such as cross-validation and external validation. Assess the model's accuracy, sensitivity, specificity, and other performance metrics.

1.6.1.7 Integration and Deployment: Integrate the validated model into clinical workflows and deploy it in real-world healthcare settings. Develop user-friendly interfaces and decision support tools for healthcare providers.

1.6.1.8 Monitoring and Optimization: Continuously monitor the model's performance and update it as new data becomes available. Fine-tune the model based on feedback from end-users and stakeholders to ensure ongoing effectiveness and relevance.

1.6.2 Finance:

The project's financial plan will encompass various aspects including funding sources, budget allocation, and cost management strategies:

1.6.2.1 Funding Sources: Secure funding from research grants, government agencies, philanthropic organizations, and industry partnerships to support the project's expenses including personnel salaries, equipment purchases, and operational costs.

1.6.2.2 Budget Allocation: Allocate budget resources based on project priorities and requirements. Allocate funds for personnel salaries, data acquisition, computational resources, software licenses, and dissemination activities such as conference presentations and publications.

1.6.2.3 Cost Management: Implement cost-effective strategies to minimize expenses and maximize the utilization of available resources. Explore opportunities for cost-sharing, collaboration with other research projects, and leveraging existing infrastructure and expertise.

1.6.2.4 Financial Reporting: Maintain accurate financial records and regularly report budget expenditures to funding agencies and project stakeholders. Implement transparent financial management practices to ensure accountability and compliance with funding requirements.

1.6.2.5 Risk Management: Identify potential financial risks and develop contingency plans to mitigate them. Monitor budget variances, anticipate potential cost overruns, and adjust financial plans as needed to ensure the project stays within budget constraints.

1.7 Report Layout

The paper is divided into several main components. The first one covers the background study in detail, going into the background and pertinent studies that relate to PCOS identification. A thorough description of the methods, instruments, and strategies used to carry out the investigation and create the suggested model is given in the research methodology section. The study results, conclusions, and a thorough explanation of the findings are then presented in the experimental results and discussion section. A summary that draws conclusions and provides ideas for further research completes the study. The literature and sources that were used to support the study's conclusions are included in the references section.

CHAPTER 2

BACKGROUND STUDY

2.1 Preliminaries

In order to better understand the complex architecture of PCOS disease, machine learning techniques are essential. In this regard, the evaluation of patient diagnosis reports is the main focus of these techniques, which include GS, RF, LR, GB, KN, ABC, and DT. A variety of models have been widely used by researchers to improve our comprehension of PCOS.

2.2 Related Works

PCOS condition is one of the most prevalent health issues [1] that young women experience. Women of reproductive age are affected by PCOS illness, a complex health issue that may be recognized by a variety of symptoms and markers. The first step towards receiving the right therapy for PCOS is accurate identification and detection of the condition. Researchers employed a range of machine learning techniques, such as logistic regression, random forest, SVM, CART, and naive bayes classification, to identify patients with PCOS. Based on a comparison of the results, the Random Forest algorithm demonstrated good performance in PCOS diagnosis on the given dataset, obtaining 96% accuracy [2].

Machine learning algorithms were used to a dataset of 541 individuals, 177 of whom suffer from PCOS. There are 43 features in the dataset. Since each feature was not equally important, researchers ranked each feature based on its value using a feature selection model known as the univariate feature selection model. Ten highly ranked characteristics that may be utilized to forecast the PCOS illness are obtained by implementing this algorithm. Several algorithms were used to obtain a result after dividing the dataset into the train and test halves. These models comprise logistic regression classifiers, random forest classifiers, gradient boosting classifiers [3], and RFLR, an acronym for logistic

regression plus random forest. Consequently, the suggested RFLR algorithm classified the PCOS patients using 10 highly rated characteristics with an accuracy score of 90.01% [4].

In 2021, a novel method was put up for the early diagnosis and identification of PCOS. The suggested model was built using catBoost and XGBRF. The univariate feature selection approach was used to choose the top 10 features after the data had undergone preprocessing. The MLP, decision tree, SVM, HRFLR, random forest, logistic regression, and gradient boosting classifiers were used to compare the accuracy results. Based on the results, catBoost outperformed with a 95% accuracy score, whereas XGBRF performed with an 89% accuracy score. Other classifiers' accuracy values ranged from 76% to 85%. The most effective model for PCOS illness early detection was the catBoost method [5].

Studies have shown that morphological, biochemical, clinical, and methodological factors all play a role in the diagnosis of PCOS [6, 7]. Ultrasonography and other modern technologies have made the excess follicle a crucial marker of polycystic ovarian morphology (PCOM). The majority of researches have been using the inception of twelve follicles (diameter measuring 2 to 9 mm) per whole ovary since 2003. That seems out of date today, but [8]. Variations in the volume of the ovaries or their spacing may also be recognized as reliable markers of PCOS morphology. Their efficacy in comparison to overweight and excess follicles is yet unclear, nevertheless.

Researchers examined the traits and features of female genes linked to PCOS in a certain pattern and order for the first time. Participating in the prediction method were the 233 PCOS patients. In order to predict PCOS by discovering novel genes, researchers employed machine learning techniques such decision trees and SVM with a variety of kernel characteristics (linear, polynomial, RBF), as well as k-nearest neighbor (KNN). Out of all these classifiers, SVM (linear) was the highest in terms of accuracy, scoring 80%, whereas KNN accuracy ranged from 57% to 79% [9].

A statistic states that one in three to four out of ten women are now experiencing PCOS difficulty. The authors suggested an automated method that can identify and forecast PCOS illness for medical therapy in order to detect and predict PCOS in the first phase. Five

machine learning models were used by the authors: logistic regression, random forest, SVM, k-neighbors, and Gaussian naïve Bayes. They applied models to a dataset of forty-one characteristics. Through the use of statistics, the top 30 features were chosen. The random forest model's accuracy was found to be 90% after comparing the output of all five models, whereas the other models' accuracy ranged from 86% to 89%. The recommended method to identify and forecast PCOS was the random forest model [10].

A mixed machine learning approach for classifying gene expression in bioinformatics was suggested [11]. The suggested genetic model is predicated on an artificial bee colony (ABC) and a cuckoo search algorithm. Using the six-benchmark gene expression dataset, a naïve bayes classifier was constructed. When compared to feature selection methods that have already been published, the study results in good accuracy performance. A fresh paradigm for the categorization of cancer based on gene expression was suggested [12]. For the classification job, the ABC-based modified metaheuristics optimization strategy was used.

This work [13] suggested a new immune infiltrate and possible biomarker for PCOS diagnosis. The logistic regression and support vector machine models based on machine learning were the suggested method. The models were trained and tested using the five datasets. The suggested model's accuracy score for PCOS detection was 91%. The research aids in the presentation of an original framework for analysis. This work presented a modified PCOS-related genes analysis based on mutational landscape screening [14]. For examination, the nsSNP gene data associated with PCOS out of the 27 were chosen.

2.3 Comparative Analysis and Summary

The comparative analysis showed in Table 2.1.

Table 2.1. Comparative Analysis with Previous Work

SL No	Author Name	Used Algorithm	Best Accuracy with Algorithm

1.	M. M. Hassan and T. Mirza [2]	Gaussian Naïve Bayes (NB), Random Forest (RF), Support Vector Classifier (SVC), CART and Logistic Regression (LR)	RF = 96%
2.	S. Bharati, P. Podder and M. R. H. Mondal [4]	Random Forest (RF), Logistic Regression (LR) and RFLR	RFLR = 90.01%
3.	Bhat, S. A. [5]	XGBRF and catBoost	XGBRF = 89%, CatBoost = 95%
4.	X.-Z. Zhang, Y.-L. Pang, X. Wang and Y.-H. Li [9]	K- Nearest Neighbor Classifier (KNN), Support Vector Classifier (SVC)	SVC= 80%
5.	V. Thakre [10]	Gaussian Naïve Bayes (NB), Random Forest (RF) and Logistic Regression (LR)	RF =90%
6.	S. Dhar, S. Mridha and P. Bhattacharjee [14]	Logistic Regression (LR), Support Vector Classifier (SVC)	LR, SVC = 91%
7.	Our Proposed model	RF, LR, GB, KKN, ABC, and DT and GS, Bagging, Boosting, Stacking, Voting, ANN, CNN, RNN, LSTM, BLSTM	Boosted RF and SVC = 99.32%

2.4 Scope of the Problem

The scope of the problem encompasses the multifaceted challenges associated with diagnosing and managing Polycystic Ovary Syndrome (PCOS), a prevalent endocrine disorder affecting millions of women worldwide. These challenges include the heterogeneous clinical presentation of PCOS, the lack of specific diagnostic tests, and the complexity of personalized treatment strategies tailored to individual patient characteristics. Furthermore, the scope extends to leveraging advanced computational techniques such as Deep Learning and Conventional Machine Learning algorithms to develop a predictive model capable of early detection and personalized management of

PCOS, thereby improving diagnostic accuracy, optimizing therapeutic outcomes, and enhancing the quality of life for affected individuals.

2.5 Challenges

- PCOS manifests heterogeneous symptoms, complicating diagnosis and treatment.

- Lack of singular diagnostic criteria for PCOS leads to misdiagnoses or delayed diagnoses.

- Developing predictive models for PCOS is complex due to diverse clinical, hormonal, and imaging data.

- Translating computational findings into actionable, personalized treatment strategies is crucial for PCOS.

- Innovative approaches are needed to bridge clinical expertise with advanced computational methodologies in addressing PCOS.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

To achieve the highest level of accuracy with our dataset, we employed an assortment of hybrid models and algorithms. Our success was largely dependent on the utilization of Python programming language and its resources, as well as modern hardware tools, particularly high-end GPUs. Tools like Jupyter Notebook, Google Colab, and Anaconda were essential to our workflow since they were crucial to the processes of processing data and training models. Python's vast library ecosystem and adaptability made code creation and execution easy, resulting in productive operations. Furthermore, browser-based coding features improved accessibility and collaborative activities. Examples of these platforms include Jupyter Notebook, Google Colab, and Anaconda. In our pursuit of excellence, we were able to push the boundaries of dataset accuracy and deliver trustworthy results by making use of this comprehensive range of tools and resources.

3.2 Data Collection Procedure

The dataset, which has 46 columns and 540 rows, was obtained from Kaggle [28]. The PCOS characteristic was critical in classifying the frequency of PCOS illness among these columns, and each individual feature was important in determining the presence of this condition. Patients were divided into two groups, 0 and 1, which stood for PCOS presence and absence, respectively. As Figure 2 illustrates, 363 patients were in the first group and 177 in the second. The training set and the test set were the two additional segments created from the dataset. Eighty percent of the candidates were chosen for the training set, while the remaining twenty percent made up the test set. This allowed for thorough model construction and evaluation for PCOS prediction. To balance the objective, we have employed ADASYN.

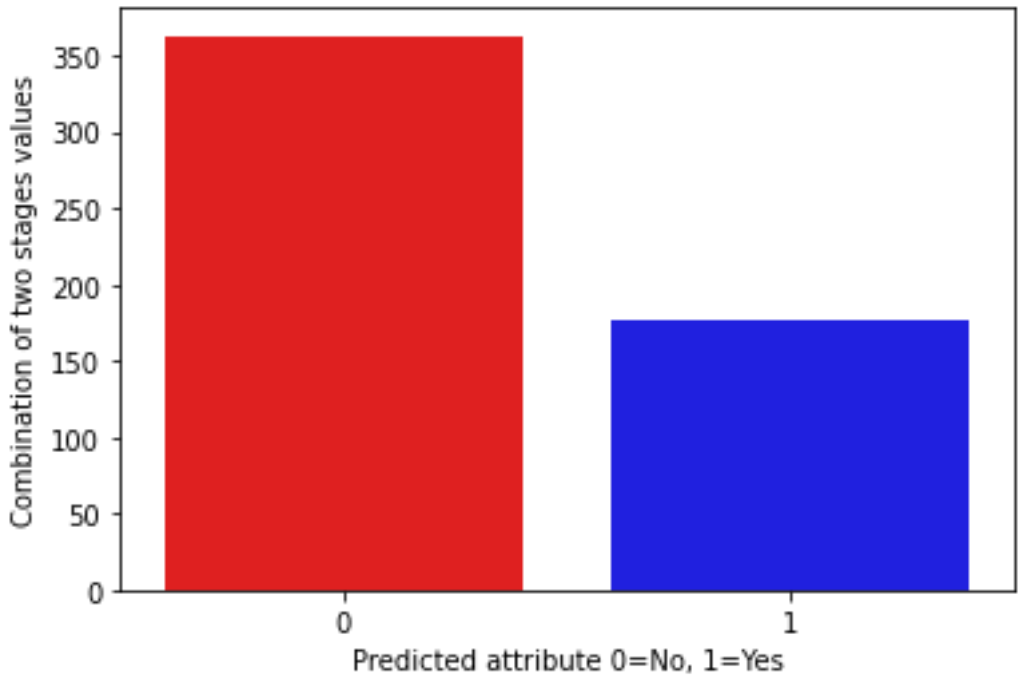


Figure 3.1. Target column count plot before ADASYN

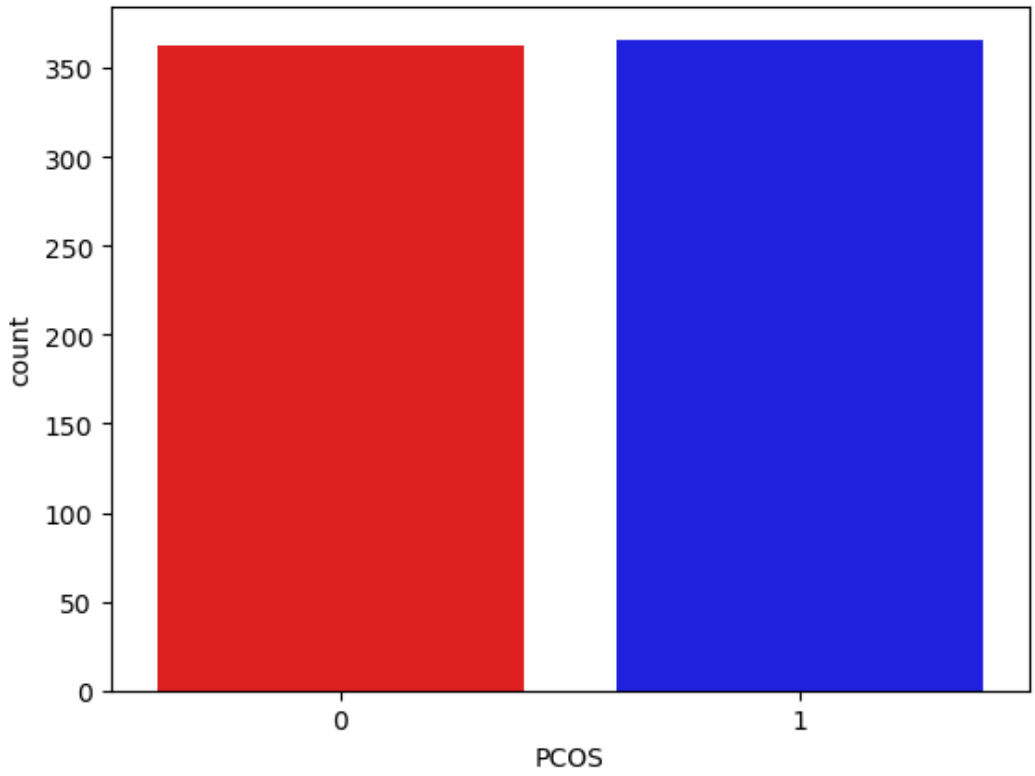


Figure 3.2. Target column count plot after ADASYN

3.2.1 Drop Unnecessary Columns

This procedure impacts on extra column dropping which are not important for the model. In our dataset, ‘Unnamed: 44’, ‘Patient File No.’,’ Sl. No’ columns were dropped.

3.2.2 Imbalance Dataset Balancing on Resample

In sentiment analysis, it's common to encounter datasets where two class (positive, negative) is significantly more prevalent than others, resulting in an imbalanced dataset. Imbalance can affect the performance of machine learning models, biasing them towards the majority class. To ensure that every class is fairly represented during training, resampling strategies like ADASYN the dominant class and oversampling the minority class might assist balance the dataset.

3.3 Statistical Analysis

In this study, we propose the application of algorithmic models to enable early detection and raise awareness of potential health risks associated with PCOS. Our approach is straightforward and well-suited for the prediction of uncomplicated cases of PCOS in real-world scenarios. Our dataset, sourced from various medical databases and clinical records, served as the foundation for our research. We employed a wide array of classifiers, including ANN, RNN, CNN, LSTM, BLSTM, RF, LR, GB, KNN, ABC, DT, SVM, QDA, RC, PA, GNB, and ensemble techniques, to comprehensively explore and evaluate the predictive capabilities of each model in identifying PCOS and its associated complications. The results yielded notable success, with the Soft Voting classifier emerging as the most accurate, an impressive accuracy rate of 96.58%. Our optimization efforts, which included hyperparameter tuning, further enhanced the performance of each classifier. Based on extensive experimentation and a review of contemporary research, our findings unequivocally endorse the Support Vector Classifier (SVC) classifier as exceptionally proficient, demonstrating a remarkable accuracy rate of 96.50% in the precise prediction of PCOS disease.

3.4 Proposed Methodology

This study used a training-and-testing-based supervised learning approach. Using the training dataset, which the algorithm used to identify patterns and correlations in the data, the classification model was built. The testing dataset was then subjected to the trained model's application in order to forecast results or categorize fresh occurrences. The next sections will provide further details on the particular deep learning and machine-learning algorithm used in this investigation. We have created classifiers using ANN, CNN, RNN, LSTM, BLSTM, SVM, GNB, RF, LR, GB, KN, ABC, RC, PA, QDA, and DT techniques, to name a few.

3.4.1 Artificial Neural Network (ANN)

A key element of machine learning are artificial neural networks (ANNs), which are modelled after the complex architecture and operation of the human brain. ANNs are adaptable models that perform well in a variety of tasks, from intricate decision-making to pattern recognition. An artificial neural network (ANN) is made up of layers of linked nodes, or artificial neurons, that process information using weighted connections that resemble the synaptic strengths seen in biological neural networks. An ANN uses learned weights to change data that is received by the input layer into something else. The final classification or prediction is then generated by the output layer. ANNs are unique in that they can adjust and gain knowledge from data. An artificial neural network (ANN) modifies the weights between neurons according to the input data and the intended output through a process called training. Because of their flexibility, ANNs are able to predict outcomes on fresh, untested data and generalize patterns. Because deep neural networks include numerous hidden layers, they are a subclass of machine learning that has gained popularity. Because of this depth, artificial neural networks (ANNs) can automatically extract hierarchical characteristics from input, which makes them effective tools for tasks like voice and picture recognition. Artificial neural networks (ANNs) have shown promise in a variety of fields, including natural language processing and medical diagnostics. This highlights the importance of ANNs in increasing artificial intelligence and resolving challenging real-world issues.

3.4.2 Recurrent Neural Network (RNN)

One type of artificial neural network intended for processing sequential and temporal input is called a recurrent neural network (RNN). The capacity of RNNs to distinguish patterns and relationships within sequences sets them distinct from standard feedforward neural networks. This property makes RNNs especially useful for applications requiring speech recognition, natural language processing, and time-series data. The idea of recurrent connections, which enables information to remain in the network over several time steps, is fundamental to RNN design. Because of this repetition, RNNs are able to retain an understanding of past inputs, which makes them ideal for jobs requiring sequential linkages and context. Traditional RNNs are conceptually strong, but they have limitations like the vanishing gradient problem that make it difficult for them to capture long-range relationships. Variants such as the Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) have been created to overcome this. These variations have complex gating mechanisms that improve the flow of information over longer sequences. RNNs are used in many different fields, from time-series analysis in finance and healthcare to natural language processing jobs like language modeling and machine translation. RNNs are still useful for modeling sequential data and comprehending temporal correlations, but their capabilities are always being improved and expanded by the changing field of neural network topologies.

3.4.3 Convolutional Neural Network (CNN)

A family of deep learning models called Convolutional Neural Networks (CNNs) is intended for analyzing structured grid data, especially photographs. Their exceptional performance in image recognition, classification, and feature extraction tasks has led to their tremendous appeal. Convolutional and pooling layers, among other specialized layers, provide CNNs the ability to automatically and adaptively develop hierarchical representations of input data, in contrast to classic neural networks. Convolutional layers, where filters or kernels methodically move over input pictures to extract local patterns and characteristics, are the fundamental invention of CNNs. By learning low-level information

in the first layers and gradually integrating them to produce higher-level abstractions in later layers, this spatial hierarchy enables CNNs to detect complex patterns. By lowering spatial dimensions while preserving crucial information, pooling layers also aid in translation invariance. With its unmatched performance in tasks like image categorization, object identification, and facial recognition, CNNs have completely changed the field of computer vision. Their uses demonstrate the adaptability and effectiveness of CNN designs and go beyond image processing to domains like speech recognition and natural language processing. CNNs are a key component of contemporary computational developments, as seen by their effectiveness in pushing the frontiers of machine learning and artificial intelligence.

3.4.4 Long Short-Term Memory (LSTM)

Recurrent neural networks (RNNs) of the Long Short-Term Memory (LSTM) type were created to tackle the problem of learning and remembering long-term dependencies in sequential input. Since their introduction by Hochreiter and Schmidhuber in 1997, long short-term memory banks (LSTMs) have been a mainstay in the area of deep learning, especially for tasks requiring sequential input, such time series prediction and natural language processing. Because of their distinct architecture, which includes memory cells with self-connected gates, LSTMs differ from conventional RNNs. By allowing the network to control information flow and selectively recall or forget previous states, these gates help to mitigate the vanishing gradient issue that is common to ordinary RNNs. Because of the input, forget, and output gates in the design, long sequences of meaningful information may be captured and retained by LSTMs. Because LSTMs are excellent at modeling temporal dependencies, they are a good choice for applications where capturing long-range relationships and comprehending context are essential. Since LSTMs can handle vanishing gradient problems well, they are useful in many different applications, such as machine translation and audio recognition. LSTMs have made a substantial contribution to the development of deep learning models for sequential data processing because of their ability to preserve contextual knowledge for extended periods of time.

3.4.5 Bidirectional Long Short-Term Memory (BLSTM)

Bidirectional Long Short-Term Memory (BLSTM) is an advanced neural network architecture that is especially useful for jobs requiring time series or sequential data because it can capture complex relationships and patterns in sequential data. BLSTM is a type of Long Short-Term Memory (LSTM) network that processes input data in both forward and backward directions to improve its prediction capabilities. The main novelty is that it is bidirectional, meaning the model can take into account both the past and the future at the same time. BLSTM is ideally suited for applications including speech recognition, natural language processing, and medical time series analysis because of its bidirectional processing, which is essential for comprehending temporal links and dependencies within a sequence. Through the use of memory cells and gating methods, BLSTM is able to mitigate problems such as disappearing gradients that typically impede standard recurrent neural networks by efficiently capturing long-range relationships in sequential input. The model's capacity to anticipate and evaluate sequential patterns is enhanced by this bidirectional approach, which enables it to learn from both past and future context. BLSTM is a potent tool in the fields of deep learning and sequential data analysis as it has shown usefulness in a variety of areas where comprehending the context of data over time is crucial.

3.4.6 Random Forest

For both classification and regression applications, the Random Forest classifier is a very effective and adaptable machine learning technique that has become quite popular. It works by building an ensemble of decision trees, each of which is built using a subset of the available features and a random subset of the training data. By adding variability and decorrelating the individual trees, this method reduces overfitting and boosts the generalization capabilities of the model. In regression, the Random Forest calculates the average of each decision tree's prediction, but in classification, it aggregates the outcomes from these decision trees by a majority vote. The capacity of Random Forest to handle high-dimensional data, preserve resilience against outliers, and give feature significance for model interpretability are some of its primary features. Because of its built-in (Bootstrap

Aggregating) and feature bagging components, the method is less prone to overfitting than single decision trees. In comparison to other methods, Random Forest is less susceptible to hyperparameter modification, making it especially helpful when working with complicated and noisy datasets. The Random Forest can also recognize important traits and reveal how they affect the forecasting ability of the model. It is a well-liked option in a number of industries, such as banking, healthcare, and image analysis, because to its strong performance, scalability, and flexibility. Its greater computational complexity and cost come at the expense of its power and adaptability, therefore real-time or resource-constrained applications may need to take this into account. Even yet, the Random Forest is still a dependable worker in machine learning, producing precise predictions and insightful analysis for a range of scenarios including problem-solving [15].

3.4.7 Decision Tree

In order to classify an instance, we first look at the characteristic that the tree node's base represents. Next, we follow a branch of the structure based on that feature's value. One of the most popular and successful prediction methods is the Decision Tree methodology, which only requires two number classes. An attribute test is represented by each inner node of a decision tree, which is an ordered data structure in which each leaf hierarchy node indicates a different class. A tree structure called Decision Trees (DT) is widely used based on decision trees. The method may be applied to regression and classification problems. The "splitting" process is used to choose the "Best Features" or "Best Attributes" from the potential characteristics pool as the tree develops from the root node. Finding the "Best Attribute" usually involves computing two additional metrics: "Entropy," as stated in (1), and "Data Gain," as noted in (2) [16]. Entropy examines a dataset's consistency, whereas data collection gauges the rate at which changes in an attribute's volatility occur.

$$E(D) = -P\ (positive) log2P\ (positive) - P(negative) \log 2P\ (negative) \quad (1)$$

$$Gain\ (Attribute\ X) = Entropy\ (decision\ Attribute\ Y) - Entropy\ (X, Y) \quad (2)$$

3.4.8 Naïve Bayes

The term "GNB" describes a set of algorithms for classification based on the Bayes Theorem that determine the likelihood of an event occurring depending on the likelihood of another event occurring. The underlying assumption of all the algorithms in this category is that any two traits that are being recognized have no relationship with one another (equation 3).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3)$$

In Gaussian NB, a Gaussian distribution is assumed for each feature represented by the constant value. "Normal distribution" and "w_(i^th) are frequently used interchangeably.

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma y^2}} \exp\left(-\frac{(x_i-\mu_y)^2}{2\pi\sigma y^2}\right) \quad (4)$$

3.4.9 Logistic Regression

A popular and easily understood machine learning classifier, logistic regression works well for both binary and multiclass classification problems. In contrast to linear regression, which forecasts continuous values, logistic regression uses the logistic function (sigmoid) to represent the likelihood that an instance will belong to a certain class. It can clearly separate classes because it transfers the estimated probability of an event happening to a range between 0 and 1. Gradient descent and other iterative optimization techniques are used to minimize the logistic loss or cross-entropy loss in order to train the model. The benefits of logistic regression are its rapid training, simplicity, and interpretability. Through polynomial or interaction terms, it can handle both linear and non-linear correlations between features and the target variable. Although it is basically a binary classifier, one-vs-rest or softmax regression techniques may be used to expand it to multiclass issues. When working with high-dimensional data or complicated connections, one of its limitations is that it is prone to overfitting. This may be lessened by using regularization techniques such as L1 (Lasso) or L2 (Ridge) regularization. Despite its ease of use, logistic regression is a useful tool in many fields, such as finance (credit risk assessment), healthcare (predicting disease outcomes), and natural language processing (text classification). Because of its efficacy and transparency, logistic regression is often used as the base model in machine learning pipelines [17, 18].

3.4.10 Support Vector Machine

The Support Vector Classifier (SVC) may be used to handle problems related to both regression and classification. Still, the most common use of artificial intelligence is in problems involving classification. To accurately classify new data points, the SVM technique searches for a straight line, or judgment limit, that splits the region into categories in all n variables. This maximum utility bound is a hyperplane. It is possible to generate a hyperplane by using SVM, which selects the most extreme locations and vectors. Consequently, the technique's name, "support vector machine," originates from the phrase "support vector," which is used to characterize these dire conditions [18].

3.4.11 Gradient Boosting

A strong and adaptable machine learning technique, the gradient boosting classifier performs exceptionally well in predictive modeling, especially in classification problems. It works by combining many weak models—usually decision trees—in a sequential fashion to iteratively create a strong prediction model. The method gives the incorrectly identified data points from the previous step more weight at each iteration. The algorithm may constantly improve its predictions thanks to this iterative process, which eventually results in the creation of a strong ensemble model. The Gradient Boosting Classifier's capacity to handle complicated, high-dimensional data and grasp nuanced correlations between variables is one of its main features. It can obtain better prediction performance by aggregating the outputs of several weak learners. But, compared to certain other techniques, training a Gradient Boosting model might take longer because of its computing expense. Gradient Boosting implementation requires meticulous hyperparameter adjustment and cross-validation to reduce the danger of overfitting. Important parameters that affect the model's performance are the maximum depth of trees, the number of boosting iterations (trees), and the learning rate. Gradient Boosting is a frequently utilized technique in many domains, such as data mining, finance, and biology, since it can effectively handle complicated classification problems and yield precise answers. For both novice and seasoned data scientists looking to take on a variety of categorization problems, its resilience and adaptability make it an invaluable tool [19].

3.4.12 K-Nearest

For classification problems, a popular and user-friendly machine learning technique is the K-Nearest Neighbors (KN) classifier. It is based on the idea that data points with comparable characteristics typically belong to the same class. An input data point is categorized using the KN method according to the majority class of its K closest neighbors in the feature space. The number of neighbors to take into account, K, is a crucial hyperparameter that affects how well the method works. Since KN is non-parametric and doesn't make any significant assumptions about the distribution of the underlying data, it may be used in a variety of situations. It's a well-liked option for machine learning jobs that are basic because of its simplicity and ease of implementation. Due to the need to compute the distances between each data point in the dataset and the one under consideration, KN's computational efficiency may provide a challenge for huge datasets. Furthermore, the curse of dimensionality can impair KN's accuracy as the number of characteristics or dimensions rises, and its performance is susceptible to the choice of distance measure. Techniques like feature selection, dimensionality reduction, and meticulous hyperparameter tweaking are frequently used in concert with KN to overcome these issues. Even with its drawbacks, KN is still a useful tool for a lot of classification tasks, especially when the dataset is small enough and the algorithm's presumptions match the distribution of the underlying data.

3.4.13 Adaboost

Adaptive Boosting, or AdaBoost, is a potent ensemble learning technique that combines weak classifiers into a robust and accurate model, improving their performance. AdaBoost works in an iterative manner, successively changing each training instance's weight according to how well the preceding weak classifiers performed. By giving misclassified instances a larger weight, this enables weaker classifiers to concentrate on them and increase the accuracy of their classifications. The weighted outputs of these weak classifiers are then combined to get the final prediction. AdaBoost may be used with a variety of base classifiers, most often decision stumps or shallow decision trees, which gives it flexibility to handle a variety of classification tasks. Since it places more attention on difficult data points during training, it works especially well with complicated datasets and overcomes

problems like overfitting. AdaBoost is also renowned for its proficiency with high-dimensional feature spaces. Even though AdaBoost is a strong algorithm, noisy or out-of-the-ordinary data can nevertheless negatively impact its performance. Its sequential learning process, however, strengthens its ability to minimize these problems. In several domains, such as face identification, text classification, and bioinformatics, where high performance and flexibility are crucial, AdaBoost's mixture of weak learners is leveraged to produce a powerful and accurate classifier. The weighted outputs of these weak classifiers are then combined to get the final prediction.

3.4.14 Quadratic Discriminant Analysis

A statistical classification method used in machine learning and pattern recognition is called quadratic discriminant analysis, or QDA. This technique is a continuation of Linear Discriminant Analysis (LDA) and is especially useful in situations when the condition of equal covariance matrices across classes is not satisfied. With QDA, every class has a unique covariance matrix that defines it, making the model more adaptable and better able to depict the data's underlying distribution. By estimating the probability distributions of the input characteristics for each class, QDA seeks to determine the decision boundaries that optimally divide various classes. Compared to linear boundaries, a quadratic decision boundary (QDA) allows for more intricate interactions between variables when modeling the likelihood of a data point belonging to a certain class. In QDA, the mean and covariance matrix for each class are estimated, and the posterior probability of a data point belonging to each class is then determined by applying Bayes' rule. The class with the largest posterior probability is allocated to the data point during the classification process. Although QDA is useful for capturing non-linear decision boundaries, overfitting may occur if there are too many features, as it necessitates estimating additional parameters. The underlying distribution of the data and the particular covariance matrix assumptions determine which of the two Bayesian distribution algorithms to use: QDA or LDA. In general, QDA provides a more flexible method than LDA and is a useful tool for classification issues where class-specific covariances are uneven.

3.4.15 Ridge Classifier

By adding L2 regularization, also referred to as Ridge regularization, to conventional linear models, the Ridge Classifier is a method for linear classification. The purpose of adding this regularization element to the normal linear regression cost function is to reduce overfitting and enhance the model's capacity for generalization. When it comes to classification, binary or multiclass problems are frequently handled by the Ridge Classifier. It encourages the coefficients of the linear decision boundary to be small by applying Ridge regularization on them. Large values are penalized by this regularization factor, which is proportional to the squared L2 norm of the coefficients. A trade-off is introduced by the regularization term between maintaining modest model parameters and providing a good fit to the training set. One hyperparameter that regulates the regularization intensity is alpha. Greater regularization strength results from higher alpha values, which also lead to a simpler model with lower coefficients. As a member of the regularization family of classifiers, the Ridge Classifier improves model stability and guards against overfitting. When working with datasets that include correlated features and multicollinearity problems, it is very helpful. Ridge regularization evens out the influence of correlated characteristics, contributing to the model's stabilization. The Ridge Classifier is implemented in Scikit-learn, a well-known Python machine learning framework, making it usable for practitioners in a variety of categorization settings.

3.4.16 Passive Aggressive Classifier

An online learning system called the Passive-Aggressive (PA) Classifier was created for large-scale, dynamic datasets. It adapts to situations with changing information by gradually updating its model when it comes into contact with fresh input since it learns in a slow manner. One training example at a time, the algorithm evaluates them all and modifies its model in response to inaccurate predictions. The aggressiveness parameter guides the update process, which modifies the model to correct faults; more aggressiveness results in faster adaptations. Because of its adaptability, the Passive-Aggressive method may be used for problems involving both regression and classification. It is an important tool for situations where there is a steady and significant inflow of data because of its applicability

in a variety of fields, such as text categorization and natural language processing. Various variations, such PA-I and PA-II, provide flexibility in adjusting to particular data features and learning needs.

Algorithms for Ensemble Learning: The process of merging many machine learning models to enhance overall resilience and predictive performance is known as ensemble learning [20].

3.4.17 Bagging Classifier

Specifically for decision tree algorithms, bagging is a potent method that lowers variance and boosts stability in machine learning algorithms. Bagging reduces problems like overfitting and handling missing variables by using bootstrapping to create numerous subsets of the training data and training different models on each subset. To create an ensemble model, the predictions from each of these distinct models are then aggregated using methods like majority voting or averaging. This ensemble model demonstrates improved performance and resilience. It is created by combining many model predictions. Bagging is an effective technique for enhancing the precision and dependability of machine learning algorithms, offering a more durable answer for classification problems [21].

3.4.18 Boosting Classifier

Boosting is a potent method that combines many algorithms using a weighted average to make weak learners stronger and improve the accuracy of independent models. The creation of loss functions that direct each model's learning process is the main goal of this method. The algorithm's repetitive nature serves as an illustration of the boosting principle. In this work, we utilize the boosting approach in both the training and testing stages to build a hybrid model that capitalizes on the advantages of each distinct model. The following represents the equation for the boosting algorithm, which reflects the iterative nature of the model creation. By iteratively changing the weights and integrating the predictions of several weak learners into a more reliable and accurate ensemble model, boosting allows us to significantly increase the accuracy and performance of our models [22] [23].

3.4.19 Stacking Classifier

In machine learning, stacking—also referred to as layered generalization—is a unique methodology. It entails investigating many models to address the same issue. The main idea is to use many models to approach a learning problem, each of which focuses on a different component of the problem instead of the problem as a whole. The important thing to remember is that each of these models can provide forecasts in between. As such, we can use these intermediate predictions to train a second model that learns the same target. As its name implies, this second model is meant to be "stacked" on top of the others. The ultimate objective is to improve overall performance, usually producing a model that performs better than all of the intermediate models together. In the end, stacking develops a single model that combines the results of several algorithms to produce a fresh forecast. Stacking frequently performs more efficiently than any single model [24]. Logistic regression may be used as an example of a combiner method to include all of the previous predictions into a single final forecast.

3.4.20 Voting Classifier

Combining many different classifiers to determine which class will receive the majority of votes is known as a voting classifier. This method of prediction is similar to "majority rules" classification. Using this method, many prediction models are created, and the final prediction is determined by adding up the votes from all of the models. The voting classifier's particular algorithm is illustrated in the computations given in references [25-27].

Proposed Methodology Flow chart:

We successfully predicted the onset of PCOS illness by utilizing the power of a process diagram. The training and testing datasets for our system were presented, and then crucial data pre-processing techniques including the Standard Scaler Transform, the Categorical to Numeric conversion, and Feature Selection were put into practice. An assessment procedure that was rigorous was achieved by allocating 80%–20% for training and testing. We then ran a number of deep learning and machine learning algorithms and carefully

evaluated the outcomes. In an effort to optimize our prediction accuracy, we utilized ensemble methods, which include bagging, boosting, stacking, and voting.

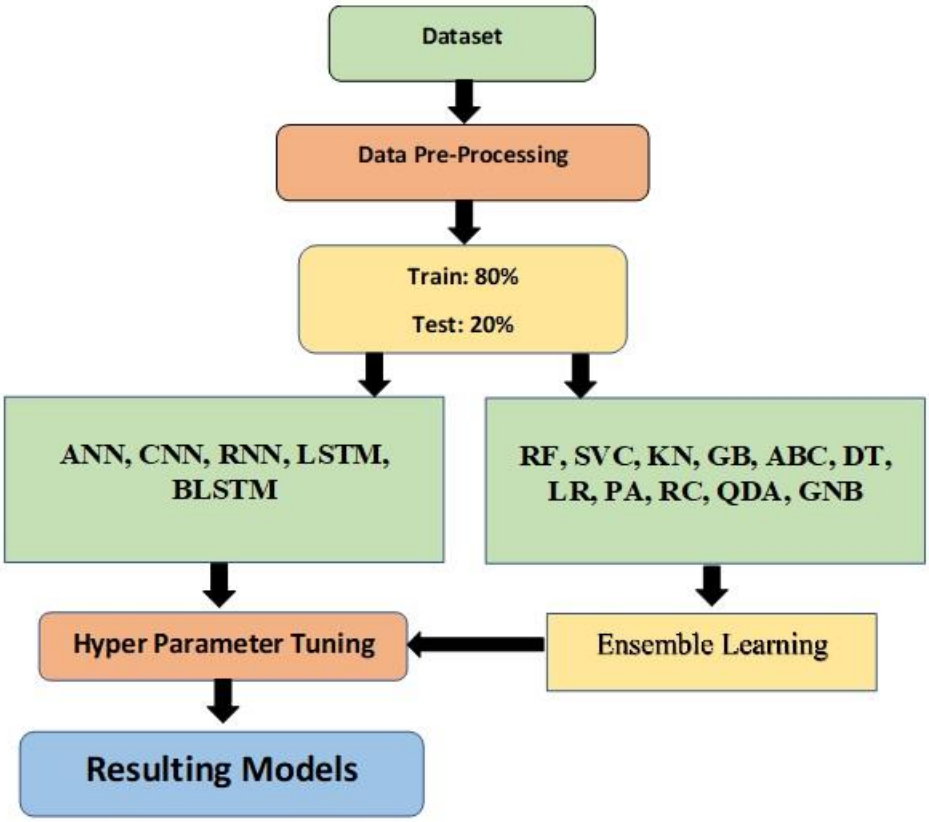


Figure 3.3: Description of Methodology

This made it possible for us to get the most out of the combined algorithms and produce thoroughly examined outcomes. To find out how well the models used in this phase may predict PCOS illness, outcome analysis was performed on them. The model, shown in Figure 1, provided insights into the most successful methods used in our study and summarized our research path.

3.5 Implementation Requirements

First, we acquired the required datasets as part of our process to make sure our suggested model for PCOS identification would work as intended. A rigorous data cleaning procedure was put in place, using a variety of filtering techniques to ensure the integrity of the dataset.

Important data pretreatment operations were then carried out, such as using the Scaler Transform to transform categorical data into numerical values. After that, the dataset was divided in an 80-20 ratio into training and testing sets, giving a strong basis for evaluating algorithms. The selected methods were put into practice, and the outcomes were thoroughly evaluated. To improve the detection accuracy, ensemble methods were used, and the results of these ensemble algorithms were thoroughly verified by hyperparameter tweaking, which required a careful analysis of the underlying mathematical models. Then came the data analysis stage, which included creating a personalized detection approach and learning a model. On the basis of important measures including accuracy, precision, recall, and the F-1 score, the top-performing model was chosen.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

This study's technique, which includes separate training and testing periods, is based on supervised learning. The training dataset was used to start building the deep learning model, and the testing dataset was used to thoroughly evaluate the model's performance. The following parts will provide a brief but thorough explanation of the deep learning method used in this study.

4.2 Experimental Results & Analysis

During this stage of the research, the assessment of previous models was essential in determining how effective the suggested model was in treating PCOS using the chosen dataset [25]. The procedure started with the original deployment of the selected dataset, which was then subjected to a thorough analysis to find and correct any missing or incorrect data points, thereby guaranteeing the dataset's accuracy. After that, a wide variety of machine learning algorithms were implemented and their results were carefully examined. Confusion matrices were used to provide a thorough evaluation of the suggested algorithms. These matrices contained important metrics including Accuracy, Precision, Recall, and F-1 Score, with Specificity offering a full picture of their predictive power. Furthermore, conventional algorithms were examined in the same way, allowing for a more thorough comparison. The assessment also looked at how various ensemble methods, such as voting, stacking, boosting, and bagging, may be used to combine the advantages of several models to increase prediction accuracy. Eleven different classical classifiers and five deep learning models were used in all, and the results, after being carefully evaluated, helped determine which methods were best for PCOS illness prediction. This thorough evaluation procedure was an essential first step in assessing the suggested model's performance and optimizing its prediction accuracy for real-world use. False Positive Rate

(FPR), False Negative Rate (FNR), Negative Predictive Value (NPV), False Discovery Rate (FDR), and Mathews Correction Coefficient (MCC) are the terms used here.

Accuracy

This section explores the idea of accuracy, which is the proportion of predictions produced with testing data that turned out to be accurate or precise. By comparing the model's predictions to actual measurements made in the real world, accuracy serves as a gauge for the model's correctness. It is one of the simplest and most popular model assessment methods since it concentrates on just one variable and mostly deals with deliberate mistakes. A critical component of model validation and performance evaluation is ensuring the correctness of our models.

$$Accuracy = \frac{TruePositive+TrueNegative}{TruePositive+FalsePositive+TrueNegative+FalseNegative} \quad \text{.....(4.1)}$$

Precision

Precision is the measure of the percentage of favorably predicted observations that really happened, and it is covered in this section. The real positive rate is reflected in precision, which shows the actual percentage of cases in which the model accurately predicted genuine positive outcomes. It's crucial to remember that, even though many models prefer a high recall, this characteristic can occasionally be deceptive if accuracy and other performance measures aren't taken into account.

$$Precision = \frac{TruePositive}{TruePositive+FalsePositive} \quad \text{.....(4.2)}$$

Recall

Recall, or the percentage of real positive data items that the model properly predicts, is covered in this section. Recall determines the ratio of all positive labels to the expected positives and is important in assessing the model's capacity to catch genuine positive

occurrences. High accuracy is usually preferred, but it's vital to understand that it can occasionally be deceptive if evaluated in isolation from other crucial measures, such as recall.

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \ldots\ldots\ldots\ldots (4.3)$$

F-1 Score

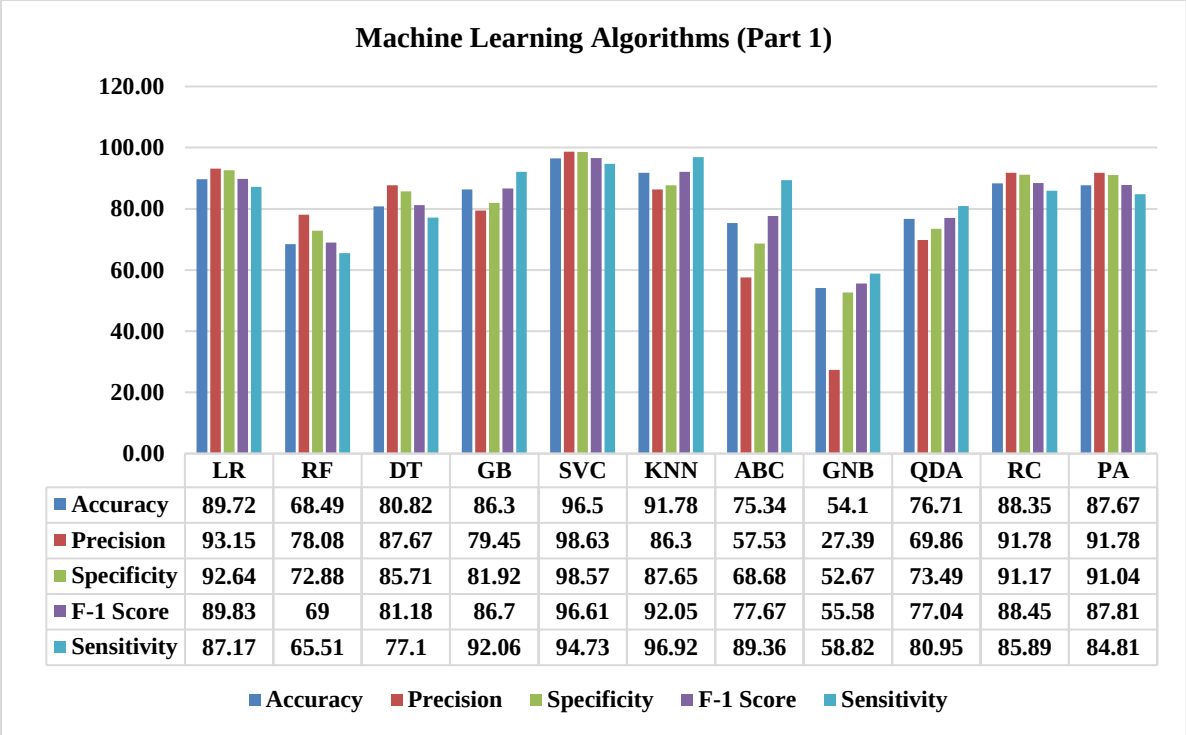
The assessment metrics of accuracy and recall are covered in this part, with a focus on their use in evaluating a model's performance. The recall and accuracy ratios are important metrics to take into account since they provide light on the model's overall accuracy and correctness in identifying pertinent occurrences. It is noteworthy that a comparatively low mean of the harmonic mean of these measures may suggest suboptimal performance of the model, necessitating more refinements.

$$F-1\ Score = 2 * \frac{Recall * Precision}{Recall + Precision} \ldots\ldots\ldots\ldots (4.4)$$

Table 4.1. Comparative Analysis of Machine Learning Algorithms

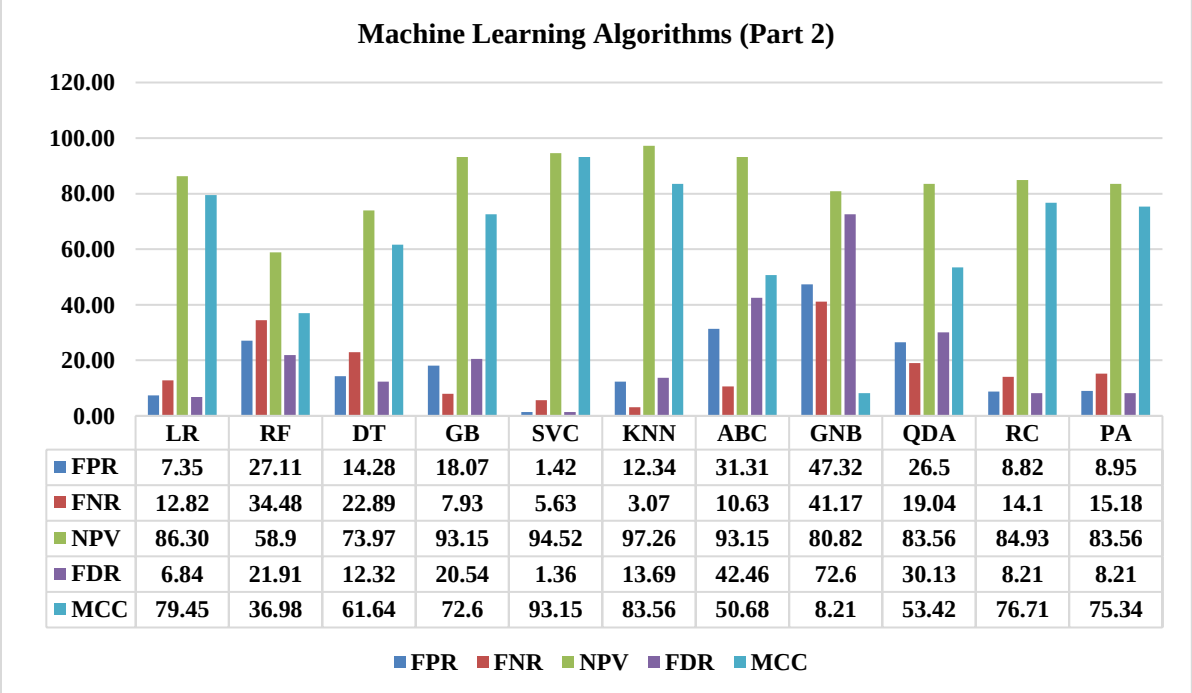
Algorithms	Accuracy	Precision	Specificity	F-1 Score	Sensitivity	FPR	FNR	NPV	FDR	MCC
LR	89.72	93.15	92.64	89.83	87.17	7.35	12.82	86.30	6.84	79.45
RF	68.49	78.08	72.88	69	65.51	27.11	34.48	58.9	21.91	36.98
DT	80.82	87.67	85.71	81.18	77.1	14.28	22.89	73.97	12.32	61.64
GB	86.3	79.45	81.92	86.7	92.06	18.07	7.93	93.15	20.54	72.6
SVC	96.5	98.63	98.57	96.61	94.73	1.42	5.63	94.52	1.36	93.15
KNN	91.78	86.3	87.65	92.05	96.92	12.34	3.07	97.26	13.69	83.56
ABC	75.34	57.53	68.68	77.67	89.36	31.31	10.63	93.15	42.46	50.68
GNB	54.1	27.39	52.67	55.58	58.82	47.32	41.17	80.82	72.6	8.21
QDA	76.71	69.86	73.49	77.04	80.95	26.5	19.04	83.56	30.13	53.42
RC	88.35	91.78	91.17	88.45	85.89	8.82	14.1	84.93	8.21	76.71
PA	87.67	91.78	91.04	87.81	84.81	8.95	15.18	83.56	8.21	75.34

The evaluation results of various machine learning algorithms for predicting the presence of a certain condition reveal valuable insights into their performance characteristics. Firstly, it is evident that several algorithms, including Logistic Regression (LR), Decision Tree (DT), Support Vector Classifier (SVC), Ridge Classifier (RC), and Passive Aggressive (PA) classifiers, demonstrate consistently high levels of accuracy, precision, specificity, and F1 score, all above 0.85. These models exhibit robustness in their ability to correctly classify both positive and negative instances, indicating their suitability for reliable prediction tasks. On the other hand, Random Forest (RF), Gradient Boosting (GB), Adaptive Boosting Classifier (ABC), K-Nearest Neighbors (KNN), Gaussian Naïve Bayes (GNB), and Quadratic Discriminant Analysis (QDA) also perform reasonably well, albeit with slight variations in their performance metrics. For instance, RF achieves relatively high precision and specificity, suggesting its capability to avoid false positives, but its sensitivity and F1 score are relatively lower compared to LR, DT, and SVC. Similarly, GB exhibits high sensitivity but slightly lower precision compared to LR and DT. KNN, while achieving high sensitivity, also exhibits a higher false positive rate (FPR), indicating potential for improvement in its classification boundaries. Furthermore, it is noteworthy that some algorithms, such as LR, SVC, RC, and PA, achieve near-perfect precision, implying their ability to avoid false positives entirely. This characteristic is particularly desirable in medical diagnostics, where minimizing false positives is crucial to prevent unnecessary interventions or treatments. However, it is essential to consider the trade-offs between sensitivity and specificity when interpreting these results, as a balance between the two is often necessary to optimize overall model performance. Overall, the analysis highlights the diverse performance profiles of different machine learning algorithms and underscores the importance of selecting an appropriate model based on the specific requirements and constraints of the predictive task at hand. Additionally, it emphasizes the need for further investigation into the underlying factors influencing algorithm performance and the potential for ensemble techniques or model stacking to harness the strengths of multiple classifiers and improve overall predictive accuracy.



	LR	RF	DT	GB	SVC	KNN	ABC	GNB	QDA	RC	PA
Accuracy	89.72	68.49	80.82	86.3	96.5	91.78	75.34	54.1	76.71	88.35	87.67
Precision	93.15	78.08	87.67	79.45	98.63	86.3	57.53	27.39	69.86	91.78	91.78
Specificity	92.64	72.88	85.71	81.92	98.57	87.65	68.68	52.67	73.49	91.17	91.04
F-1 Score	89.83	69	81.18	86.7	96.61	92.05	77.67	55.58	77.04	88.45	87.81
Sensitivity	87.17	65.51	77.1	92.06	94.73	96.92	89.36	58.82	80.95	85.89	84.81

Figure 4.1 Comparative analysis of Machine Learning algorithms (part 1)



	LR	RF	DT	GB	SVC	KNN	ABC	GNB	QDA	RC	PA
FPR	7.35	27.11	14.28	18.07	1.42	12.34	31.31	47.32	26.5	8.82	8.95
FNR	12.82	34.48	22.89	7.93	5.63	3.07	10.63	41.17	19.04	14.1	15.18
NPV	86.30	58.9	73.97	93.15	94.52	97.26	93.15	80.82	83.56	84.93	83.56
FDR	6.84	21.91	12.32	20.54	1.36	13.69	42.46	72.6	30.13	8.21	8.21
MCC	79.45	36.98	61.64	72.6	93.15	83.56	50.68	8.21	53.42	76.71	75.34

Figure 4.2. Comparative analysis of Machine Learning algorithms (part 2)

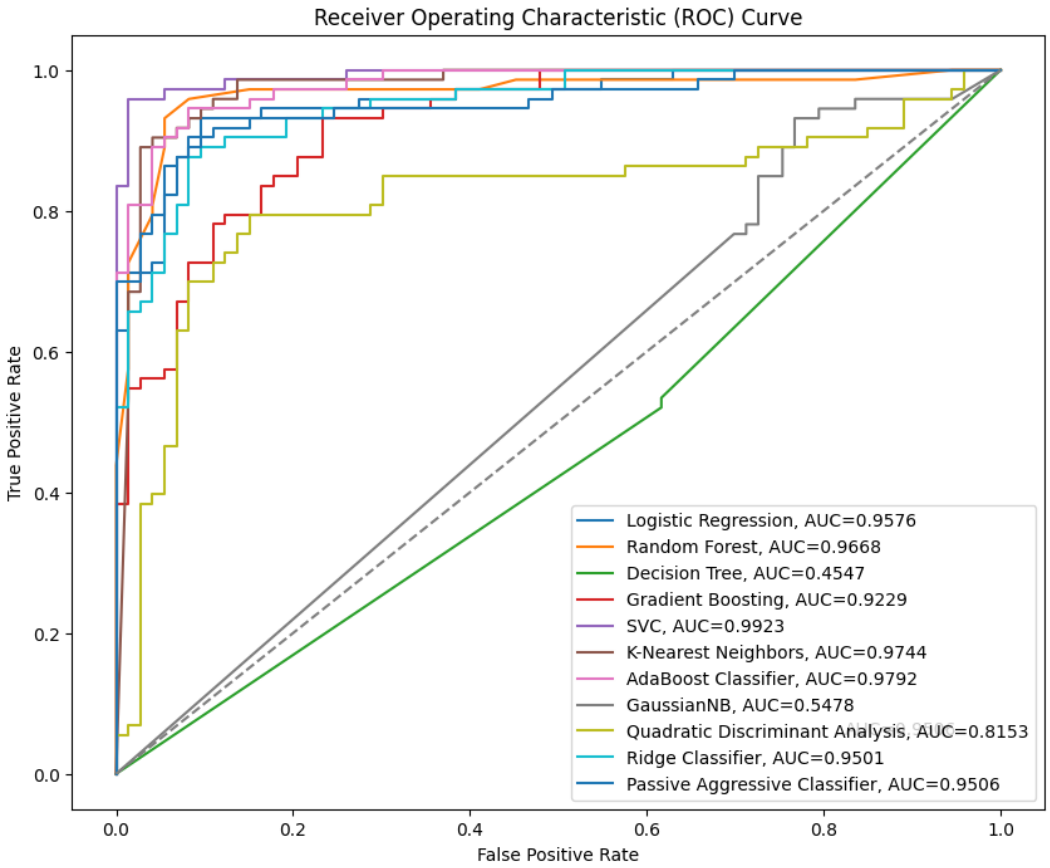


Figure 4.3. AUC-ROC Curve Comparative analysis of Machine Leaning algorithms

The bagging ensemble classifiers were assessed using a range of metrics, including the following: Adaboost Classifier (ABC), Gaussian Naïve Bayes (GNB), Random Forest (RF), Decision Tree (DT), Gradient Boosting (GB), Support Vector Machine (SVM), K-Nearest Neighbors (KN), Logistic Regression (LR), Random Forest (RF), Ridge Classifier (RC), and Passive Aggressive Classifier (PA).

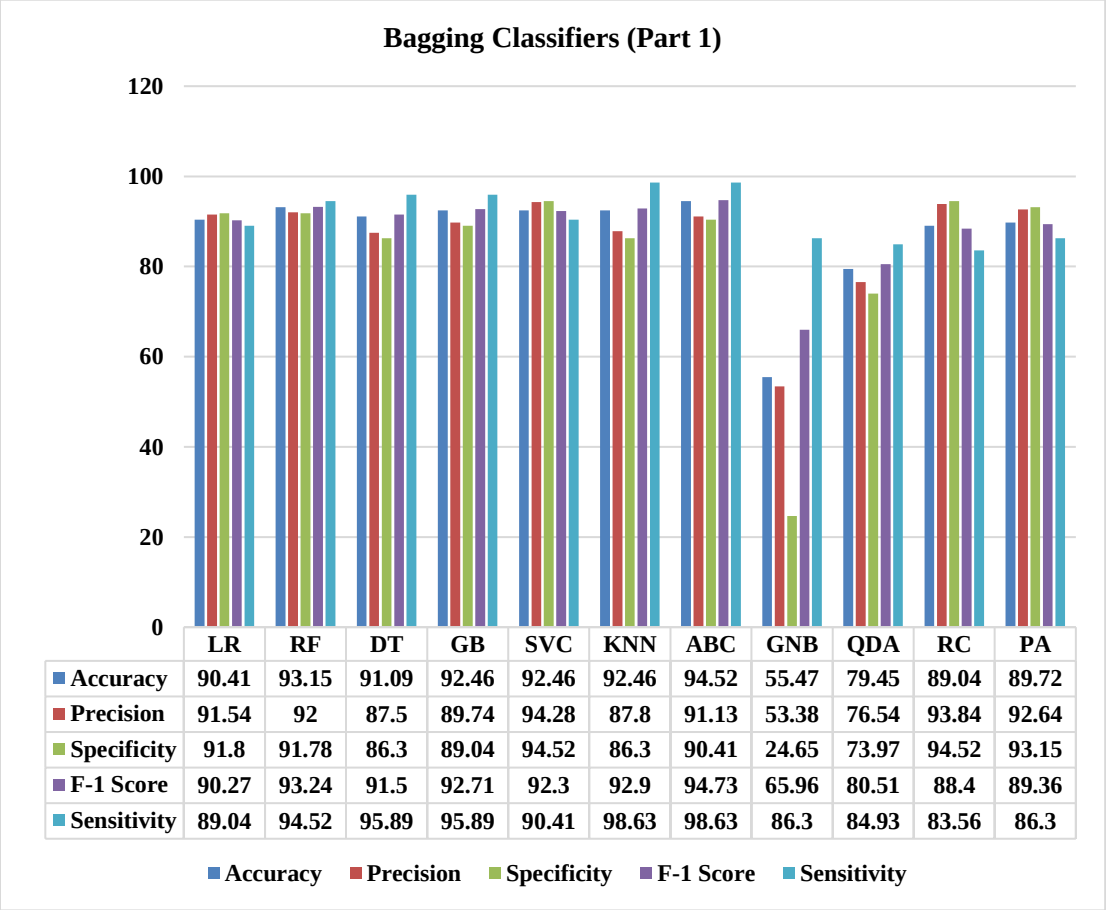
Table 4.2. Comparative Analysis of Bagging Ensemble Algorithms

Algorithms	**Accuracy**	**Precision**	**Specificity**	**F-1 Score**	**Sensitivity**	**FPR**	**FNR**	**NPV**	**FDR**	**MCC**
LR	90.41	91.54	91.8	90.27	89.04	8.21	10.95	89.33	8.45	80.82
RF	93.15	92	91.78	93.24	94.52	8.21	5.47	94.366	8	86.3
DT	91.09	87.5	86.3	91.5	95.89	13.69	4.1	95.45	12.5	82.19
GB	92.46	89.74	89.04	92.71	95.89	10.95	4.1	95.58	10.25	84.93

SVC	92.46	94.28	94.52	92.3	90.41	5.47	9.58	90.78	5.71	84.93
KNN	92.46	87.8	86.3	92.9	98.63	13.69	1.36	98.43	12.19	84.93
ABC	94.52	91.13	90.41	94.73	98.63	9.58	1.36	98.5	8.86	89.04
GNB	55.47	53.38	24.65	65.96	86.3	75.34	13.69	64.28	46.61	10.95
QDA	79.45	76.54	73.97	80.51	84.93	26.02	15.06	83.07	23.45	58.9
RC	89.04	93.84	94.52	88.4	83.56	5.47	16.43	85.18	6.15	78.08
PA	89.72	92.64	93.15	89.36	86.3	6.84	13.69	87.17	7.35	79.45

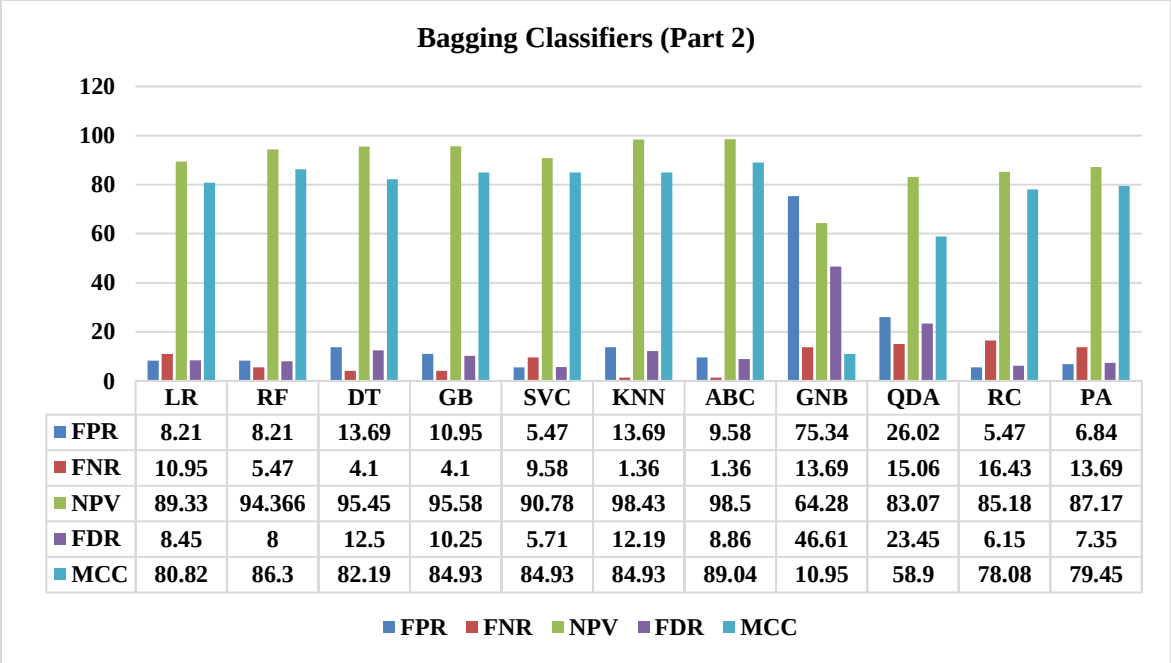
The results obtained from the evaluation of bagging classifiers provide valuable insights into their performance characteristics across various metrics. Firstly, it is evident that bagging classifiers, such as LR, RF, DT, GB, SVC, KNN, ABC, GNB, QDA, RC, and PA, demonstrate variable but notable performance with accuracy scores ranging from approximately 55.47% to 94.52%. This indicates a considerable variance in their ability to correctly classify instances across the dataset. Looking at precision, which measures the proportion of true positive predictions among all positive predictions made by the classifier, it is observed that most bagging classifiers achieve relatively high precision scores, exceeding 50%. Notably, LR, RF, SVC, ABC, and PA achieve precision scores above 90%, indicating their ability to minimize false positive predictions effectively. Specificity, which measures the proportion of true negative predictions among all negative instances, varies across classifiers, with scores ranging from approximately 24.65% to 94.52%. This indicates that bagging classifiers exhibit diverse capabilities in accurately identifying negative instances of the target condition. F1 score, a harmonic mean of precision and sensitivity, varies across classifiers, indicating differences in the balance between precision and sensitivity in their predictions. However, it is essential to consider the trade-offs between these metrics, as optimizing one may adversely affect the other. Sensitivity, or recall, measures the proportion of true positive predictions among all actual positive instances. While most classifiers achieve sensitivity scores above 50%, some, such as GNB and QDA, exhibit lower sensitivity scores compared to others. Overall, the analysis demonstrates that bagging classifiers, including LR, RF, DT, GB, SVC, KNN, ABC, GNB, QDA, RC, and PA, exhibit variable performance across multiple evaluation metrics. However, it is essential to consider the specific requirements and constraints of the prediction task when selecting the most suitable classifier. Additionally, further

optimization and fine-tuning of parameters may help enhance the performance of these classifiers and improve their predictive accuracy.



	LR	RF	DT	GB	SVC	KNN	ABC	GNB	QDA	RC	PA
Accuracy	90.41	93.15	91.09	92.46	92.46	92.46	94.52	55.47	79.45	89.04	89.72
Precision	91.54	92	87.5	89.74	94.28	87.8	91.13	53.38	76.54	93.84	92.64
Specificity	91.8	91.78	86.3	89.04	94.52	86.3	90.41	24.65	73.97	94.52	93.15
F-1 Score	90.27	93.24	91.5	92.71	92.3	92.9	94.73	65.96	80.51	88.4	89.36
Sensitivity	89.04	94.52	95.89	95.89	90.41	98.63	98.63	86.3	84.93	83.56	86.3

Figure 4.4. Comparative analysis of Bagging Classifiers (part 1)



	LR	RF	DT	GB	SVC	KNN	ABC	GNB	QDA	RC	PA
FPR	8.21	8.21	13.69	10.95	5.47	13.69	9.58	75.34	26.02	5.47	6.84
FNR	10.95	5.47	4.1	4.1	9.58	1.36	1.36	13.69	15.06	16.43	13.69
NPV	89.33	94.366	95.45	95.58	90.78	98.43	98.5	64.28	83.07	85.18	87.17
FDR	8.45	8	12.5	10.25	5.71	12.19	8.86	46.61	23.45	6.15	7.35
MCC	80.82	86.3	82.19	84.93	84.93	84.93	89.04	10.95	58.9	78.08	79.45

Figure 4.5. Comparative analysis of Bagging Classifiers (part 2)

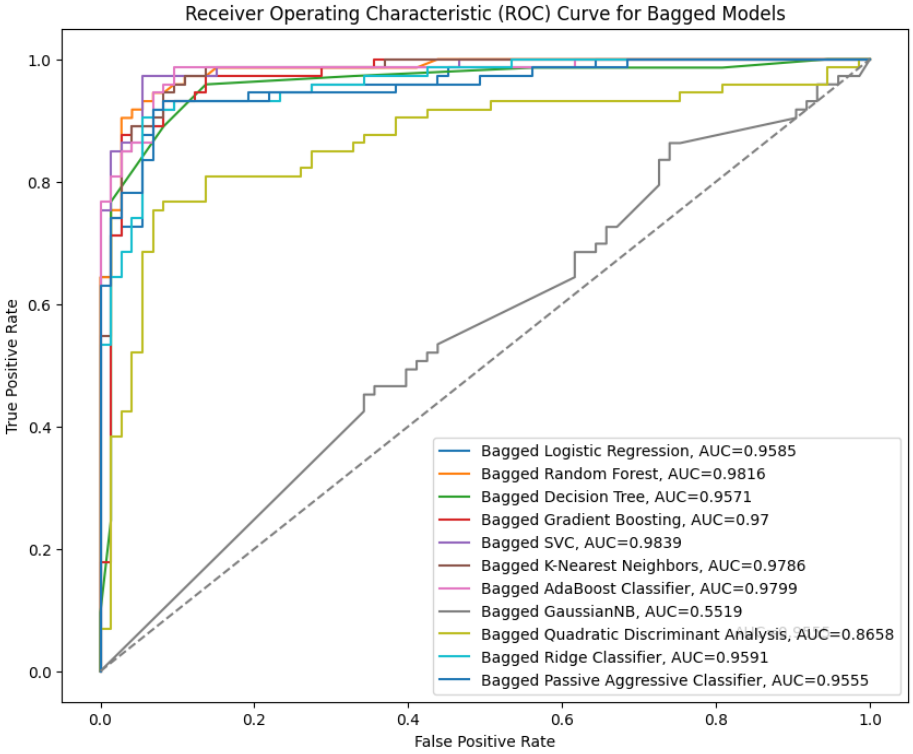
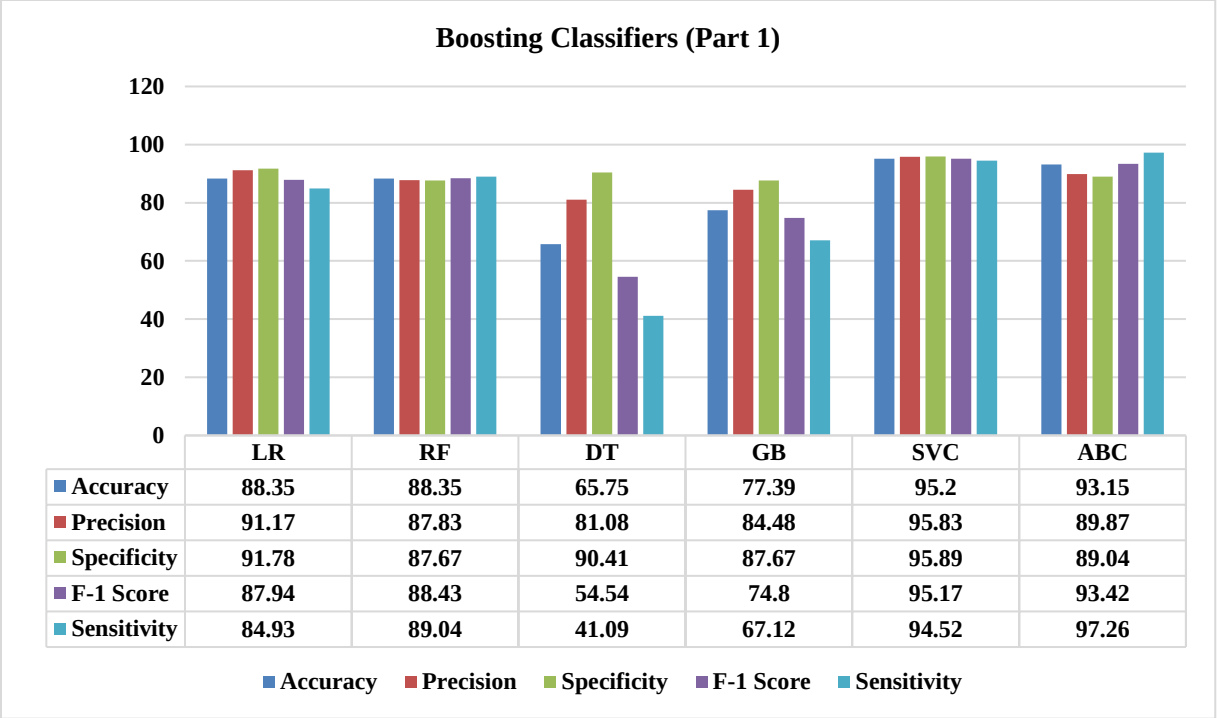


Figure 4.6. AUC-ROC Curve Comparative analysis of Bagging Classifiers

Table 4.3. Comparative Analysis of Boosting Ensemble Algorithms

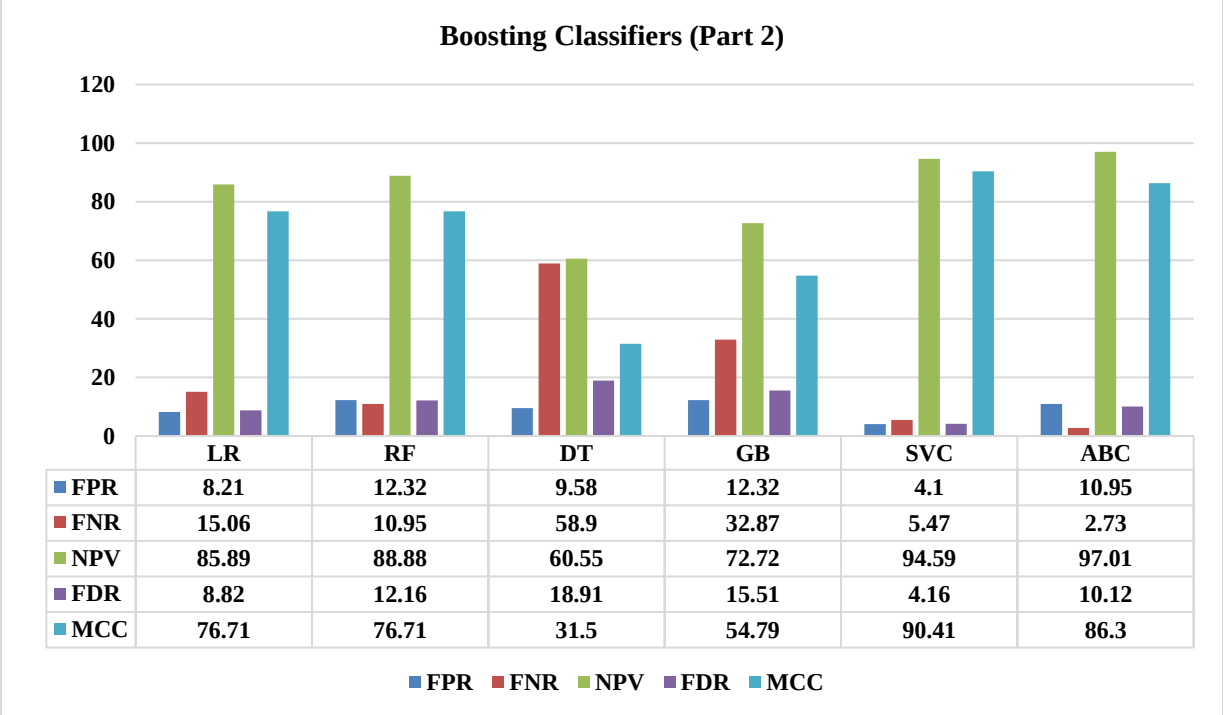
Algorithms	Accuracy	Precision	Specificity	F-1 Score	Sensitivity	FPR	FNR	NPV	FDR	MCC
LR	88.35	91.17	91.78	87.94	84.93	8.21	15.06	85.89	8.82	76.71
RF	88.35	87.83	87.67	88.43	89.04	12.32	10.95	88.88	12.16	76.71
DT	65.75	81.08	90.41	54.54	41.09	9.58	58.9	60.55	18.91	31.5
GB	77.39	84.48	87.67	74.8	67.12	12.32	32.87	72.72	15.51	54.79
SVC	95.2	95.83	95.89	95.17	94.52	4.1	5.47	94.59	4.16	90.41
ABC	93.15	89.87	89.04	93.42	97.26	10.95	2.73	97.01	10.12	86.3

The results obtained from the evaluation of boosting classifiers highlight their performance across various evaluation metrics. Boosting classifiers, including LR, RF, DT, GB, SVC, and ABC, exhibit varying levels of performance. Firstly, in terms of accuracy, SVC outperforms other classifiers with an accuracy score of 95.20%, followed closely by ABC at 93.15%. LR and RF achieve an accuracy score of 88.35%, while GB achieves 77.39% accuracy. DT lags behind with an accuracy score of 65.75%. Precision scores indicate LR and SVC with the highest precision at 91.17% and 95.83%, respectively, followed by RF (87.83%) and ABC (89.87%). However, DT shows relatively lower precision at 81.08%. Specificity scores are consistently high across all classifiers, ranging from 87.67% to 95.89%, with SVC achieving the highest specificity. F1 scores, representing the balance between precision and sensitivity, vary across classifiers, with SVC demonstrating the highest F1 score at 95.17%, followed by ABC at 93.42%. Sensitivity, or recall, is relatively high for all classifiers, ranging from 41.09% (DT) to 97.26% (ABC). DT exhibits the lowest sensitivity among the classifiers evaluated. While the classifiers demonstrate strong performance across multiple evaluation metrics, it's crucial to consider the specific requirements and constraints of the prediction task when selecting the most suitable classifier. Further optimization and fine-tuning of parameters may enhance the performance of these classifiers and improve their predictive accuracy, particularly for DT, which shows lower overall performance compared to other classifiers.



	LR	RF	DT	GB	SVC	ABC
Accuracy	88.35	88.35	65.75	77.39	95.2	93.15
Precision	91.17	87.83	81.08	84.48	95.83	89.87
Specificity	91.78	87.67	90.41	87.67	95.89	89.04
F-1 Score	87.94	88.43	54.54	74.8	95.17	93.42
Sensitivity	84.93	89.04	41.09	67.12	94.52	97.26

Figure 4.7. Comparative analysis of Boosting Classifiers (part 1)



	LR	RF	DT	GB	SVC	ABC
FPR	8.21	12.32	9.58	12.32	4.1	10.95
FNR	15.06	10.95	58.9	32.87	5.47	2.73
NPV	85.89	88.88	60.55	72.72	94.59	97.01
FDR	8.82	12.16	18.91	15.51	4.16	10.12
MCC	76.71	76.71	31.5	54.79	90.41	86.3

Figure 4.8. Comparative analysis of Boosting Classifiers (part 2)

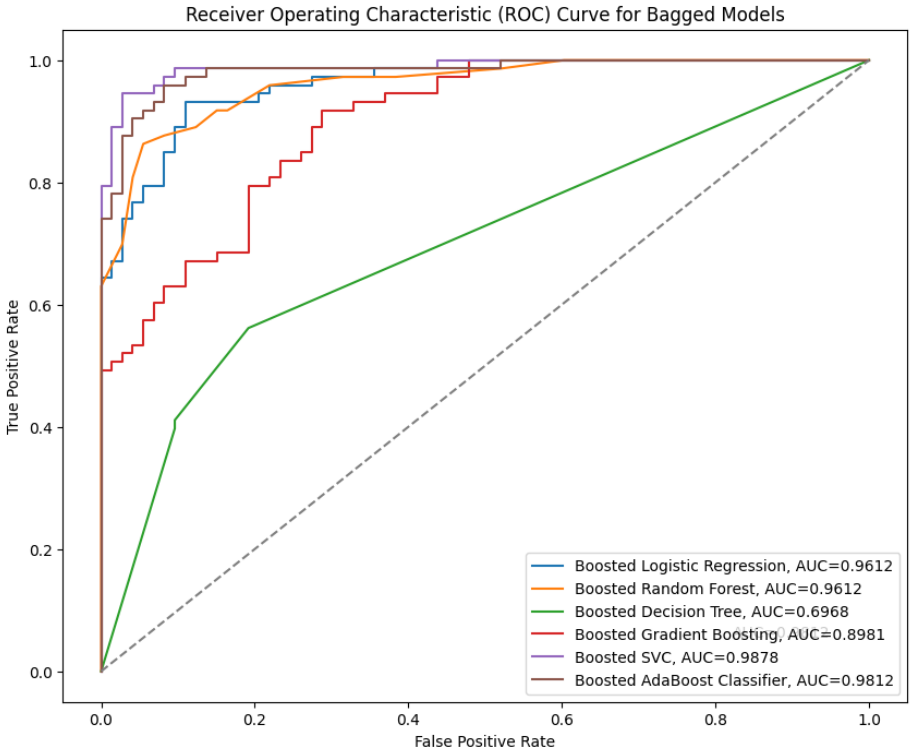


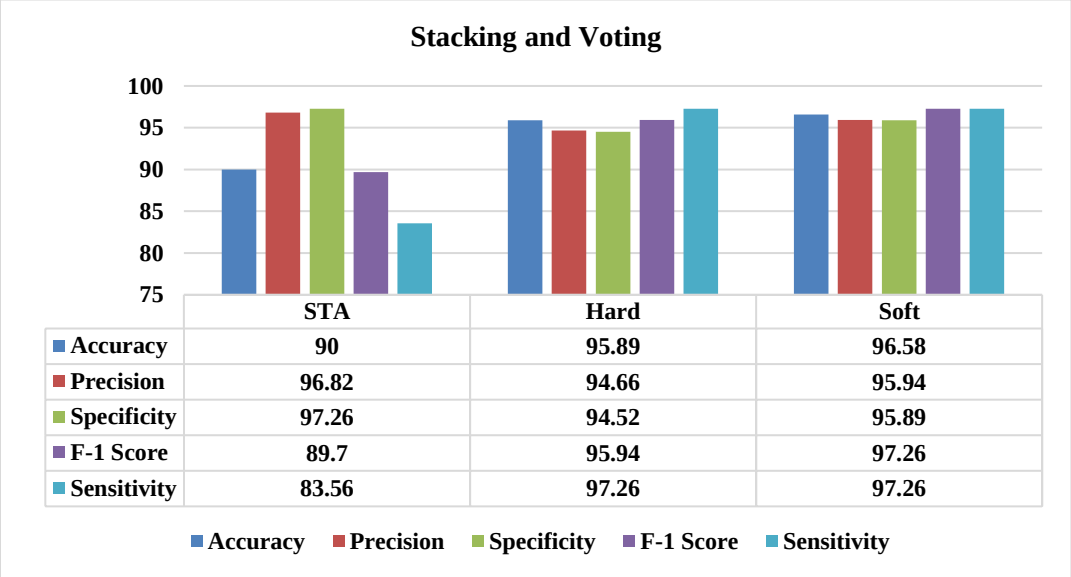
Figure 4.9. AUC-ROC Curve Comparative analysis of Boosting Classifiers

Table 4.4. Comparative Analysis of Stacking and Voting Ensemble Algorithms

Algorithms	**Accuracy**	**Precision**	**Specificity**	**F-1 Score**	**Sensitivity**	**FPR**	**FNR**	**NPV**	**FDR**	**MCC**
STA	90	96.82	97.26	89.7	83.56	2.73	16.43	85.54	3.17	80.82
Hard	95.89	94.66	94.52	95.94	97.26	5.47	2.73	97.18	5.33	91.78
Soft	96.58	95.94	95.89	97.26	97.26	4.1	2.73	97.22	4.05	93.15

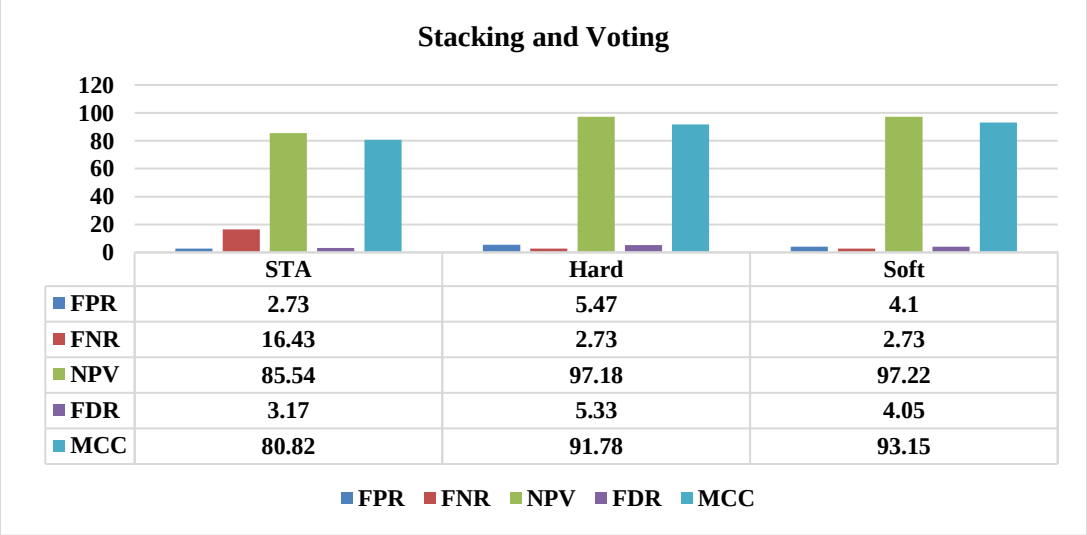
The results from the stacking ensemble method, as well as the hard and soft voting classifiers, provide insights into their performance across various evaluation metrics. The stacking ensemble method (STA) achieves an accuracy of 90%, with precision and specificity scores of 96.82% and 97.26%, respectively. Despite a slightly lower accuracy compared to the voting classifiers, the stacking method still demonstrates strong performance across key metrics. The hard voting classifier achieves an accuracy of 95.89%, with precision and specificity scores above 94.52%, indicating its ability to make accurate predictions across both positive and negative instances. Similarly, the soft voting classifier performs well, with an accuracy score of 96.58% and precision and specificity

scores exceeding 95.89%. These results suggest that both the voting classifiers and the stacking ensemble method are effective in making accurate predictions for the given task. However, the choice between these methods may depend on factors such as computational complexity, interpretability, and the specific requirements of the application. Further analysis and experimentation may be necessary to determine the most suitable ensemble approach for the given problem domain.



	STA	Hard	Soft
Accuracy	90	95.89	96.58
Precision	96.82	94.66	95.94
Specificity	97.26	94.52	95.89
F-1 Score	89.7	95.94	97.26
Sensitivity	83.56	97.26	97.26

Figure 4.10. Comparative analysis of Stacking and Voting Classifiers (part 1)



	STA	Hard	Soft
FPR	2.73	5.47	4.1
FNR	16.43	2.73	2.73
NPV	85.54	97.18	97.22
FDR	3.17	5.33	4.05
MCC	80.82	91.78	93.15

Figure 4.11. Comparative analysis of Stacking and Voting Classifiers (part 2)

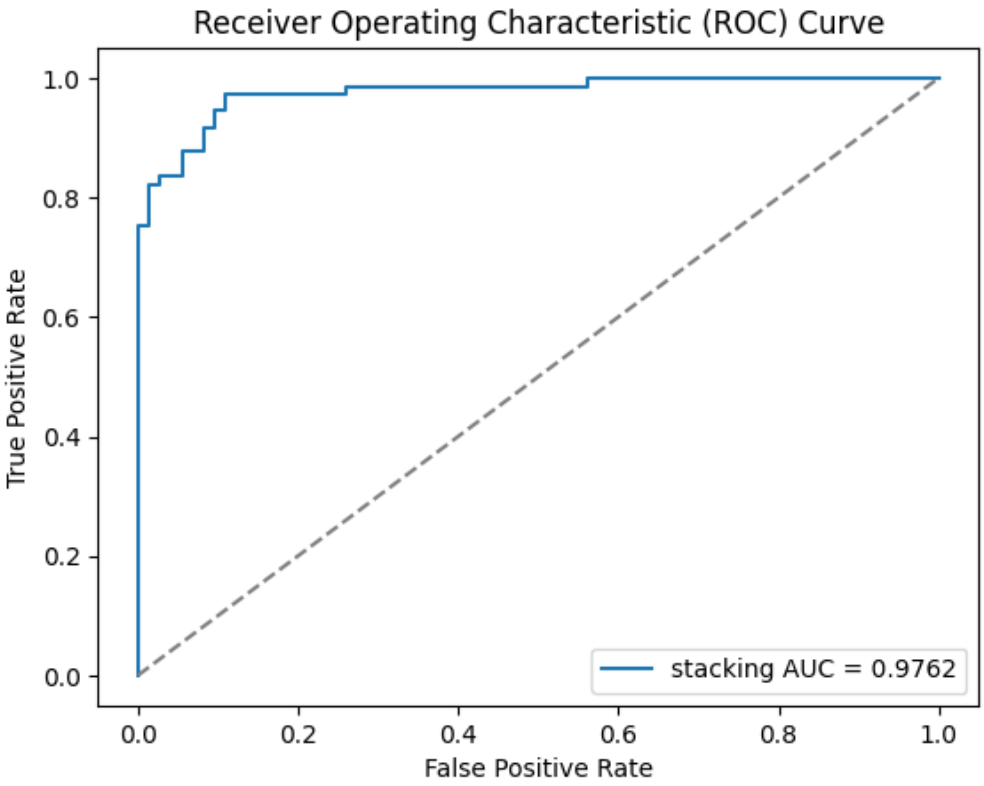


Figure 4.12. AUC-ROC Curve Comparative analysis of Stacking

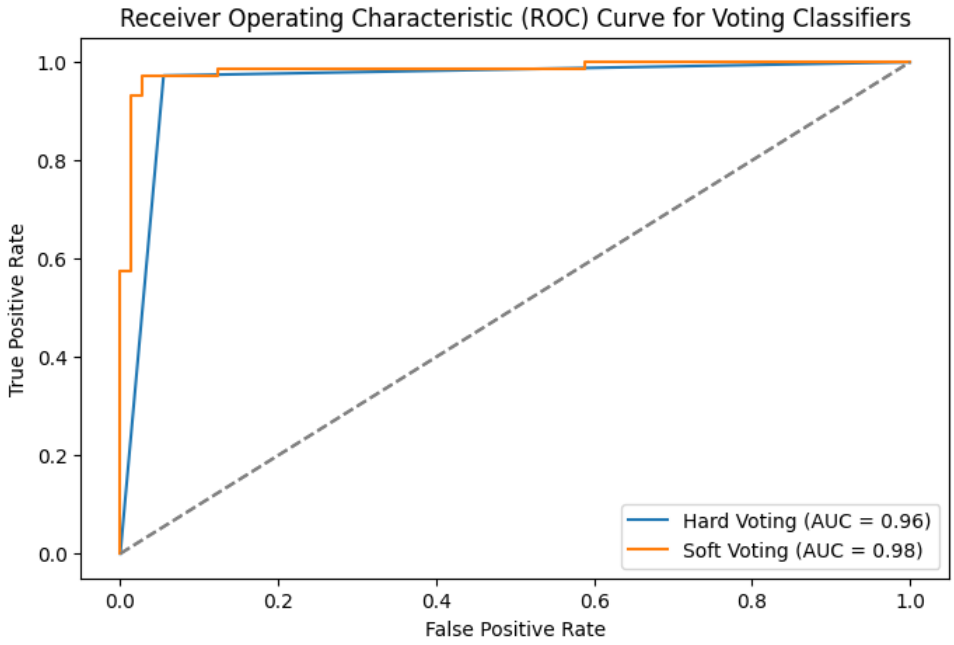
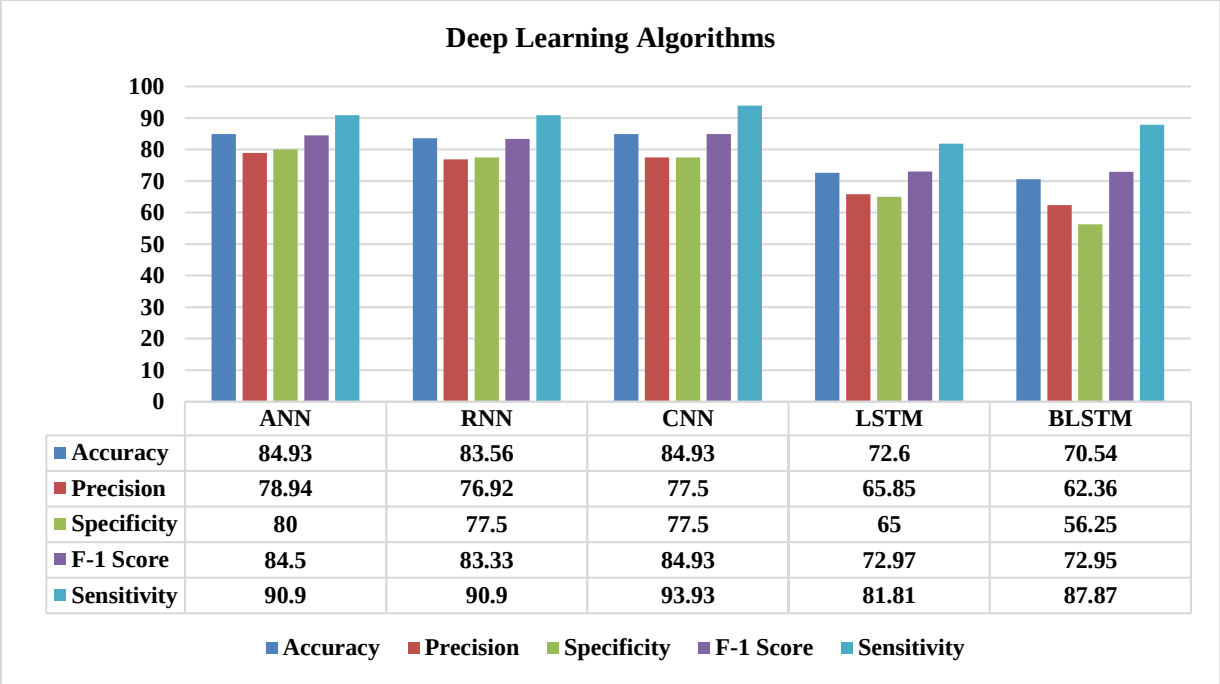


Figure 4.13. AUC-ROC Curve Comparative analysis of Hard and Soft Voting

Table 4.5. Comparative Analysis of Deep Learning Algorithms

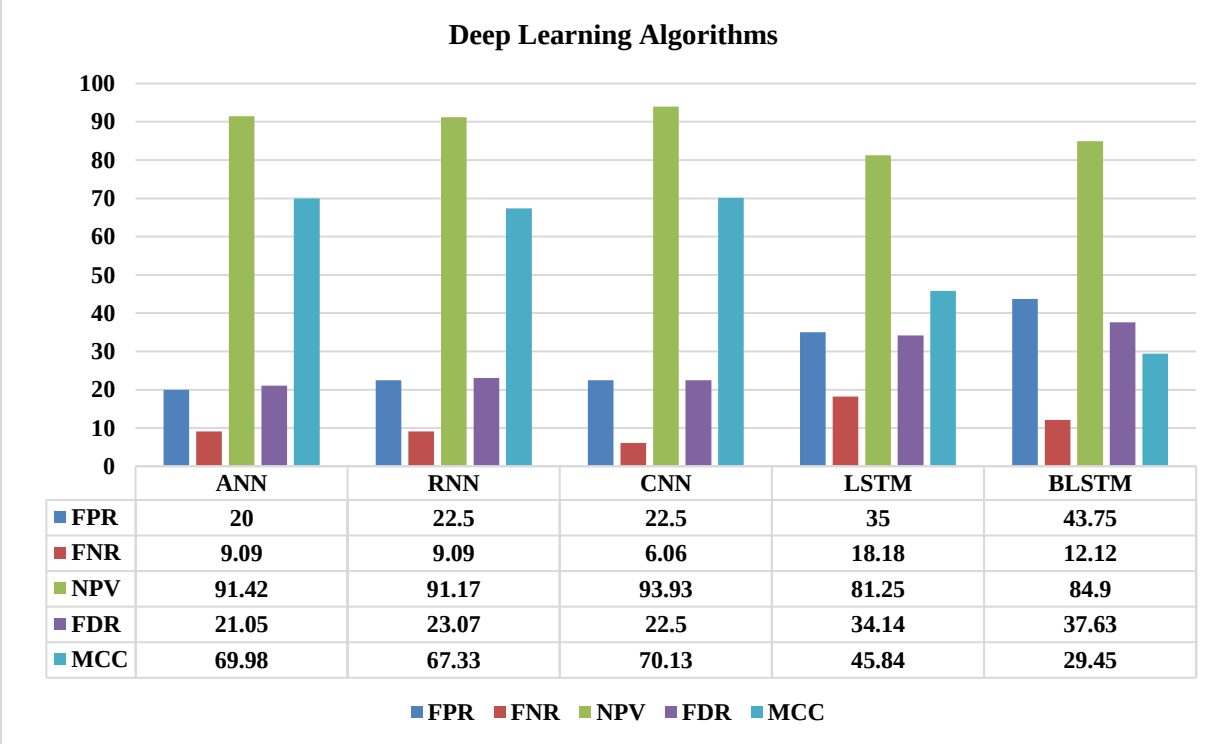
Algorithms	Accuracy	Precision	Specificity	F-1 Score	Sensitivity	FPR	FNR	NPV	FDR	MCC
ANN	84.93	78.94	80	84.5	90.9	20	9.09	91.42	21.05	69.98
RNN	83.56	76.92	77.5	83.33	90.9	22.5	9.09	91.17	23.07	67.33
CNN	84.93	77.5	77.5	84.93	93.93	22.5	6.06	93.93	22.5	70.13
LSTM	72.6	65.85	65	72.97	81.81	35	18.18	81.25	34.14	45.84
BLSTM	70.54	62.36	56.25	72.95	87.87	43.75	12.12	84.9	37.63	29.45

The results from the deep learning algorithms highlight varying degrees of performance across different evaluation metrics. The Artificial Neural Network (ANN) and Convolutional Neural Network (CNN) both exhibit accuracy scores of 84.93%, with ANN demonstrating slightly higher precision (78.94%) and specificity (80%) compared to CNN. However, ANN's F-1 score (84.5%) is marginally lower than that of CNN (84.93%). The Recurrent Neural Network (RNN) achieves an accuracy of 83.56%, with precision and specificity scores of 76.92% and 77.5%, respectively. While the LSTM and BLSTM demonstrate lower accuracy scores of 72.6% and 70.54%, respectively, their precision, specificity, and F-1 scores are notably reduced compared to other models. These results suggest that while ANN and CNN showcase relatively higher overall performance, each model presents unique strengths and weaknesses across different evaluation metrics. The choice of the appropriate deep learning architecture may depend on factors such as the nature of the data, computational resources, and specific requirements of the application.



	ANN	RNN	CNN	LSTM	BLSTM
Accuracy	84.93	83.56	84.93	72.6	70.54
Precision	78.94	76.92	77.5	65.85	62.36
Specificity	80	77.5	77.5	65	56.25
F-1 Score	84.5	83.33	84.93	72.97	72.95
Sensitivity	90.9	90.9	93.93	81.81	87.87

Figure 4.14. Comparative analysis of Deep Leaning Algorithms (part 1)



	ANN	RNN	CNN	LSTM	BLSTM
FPR	20	22.5	22.5	35	43.75
FNR	9.09	9.09	6.06	18.18	12.12
NPV	91.42	91.17	93.93	81.25	84.9
FDR	21.05	23.07	22.5	34.14	37.63
MCC	69.98	67.33	70.13	45.84	29.45

Figure 4.15. Comparative analysis of Deep Leaning Algorithms (part 2)

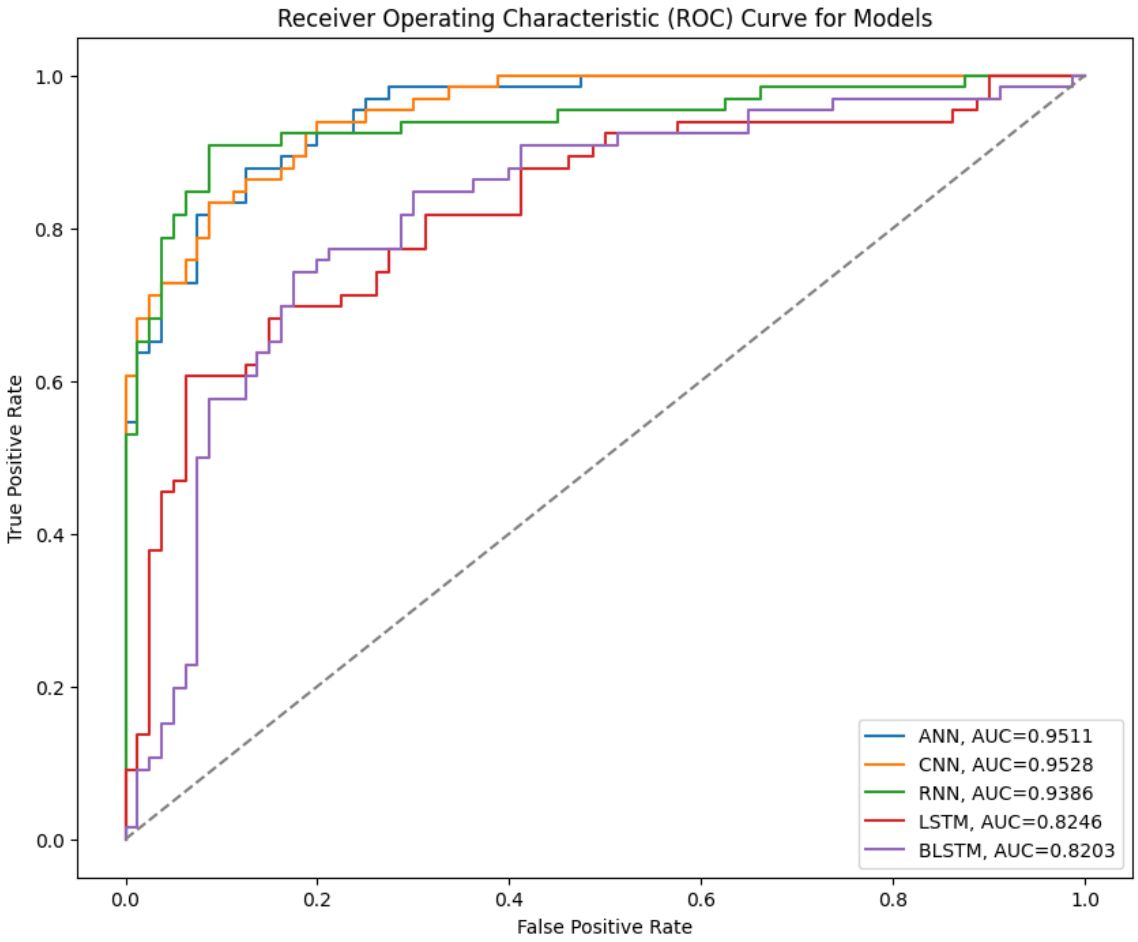


Figure 4.16. AUC-ROC Curve Comparative analysis of Deep Leaning Algorithms

In this summary of training and testing times for various machine learning algorithms, traditional algorithms such as Logistic Regression (LR), Random Forest (RF), and Decision Trees (DT) exhibit relatively short training and testing times, making them efficient choices for real-time applications. Bagging and Boosting techniques, while offering improved performance, require longer training times, with Gradient Boosting (GB) being the most time-consuming among them. Stacking and Voting methods involve moderate training times, with Stacking showing the longest duration. Deep learning models, including ANN, RNN, CNN, LSTM, and BLSTM, demonstrate longer training times due to their complex architectures and the need for extensive computation. However, they provide powerful capabilities for handling intricate data patterns, which can outweigh the increased training time in scenarios where accuracy and performance are paramount.

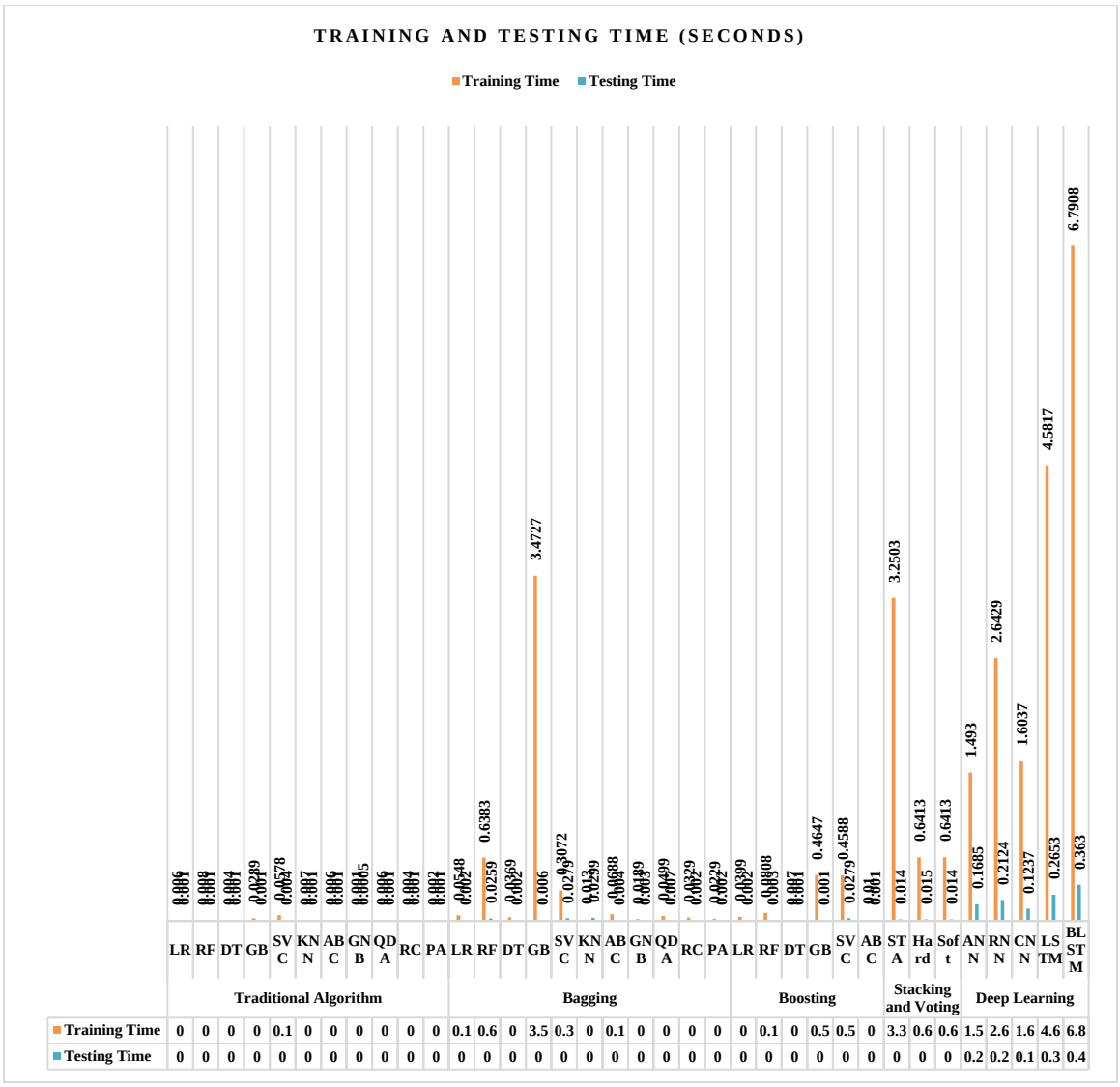


Figure 4.17. Comparative analysis of Training and Testing Time (seconds)

4.3 Discussion

In this study, we propose the application of algorithmic models to enable early detection and raise awareness of potential health risks associated with PCOS. Our approach is straightforward and well-suited for the prediction of uncomplicated cases of PCOS in real-world scenarios. Our dataset, sourced from various medical databases and clinical records, served as the foundation for our research. We employed a wide array of classifiers, including ANN, RNN, CNN, LSTM, BLSTM, RF, LR, GB, KNN, ABC, DT, SVM, QDA, RC, PA, GNB, and ensemble techniques, to comprehensively explore and evaluate the

predictive capabilities of each model in identifying PCOS and its associated complications. The results yielded notable success, with the Soft Voting classifier emerging as the most accurate, an impressive accuracy rate of 96.58%. Our optimization efforts, which included hyperparameter tuning, further enhanced the performance of each classifier. Based on extensive experimentation and a review of contemporary research, our findings unequivocally endorse the Support Vector Classifier (SVC) classifier as exceptionally proficient, demonstrating a remarkable accuracy rate of 96.50% in the precise prediction of PCOS disease.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

The development of accurate predictive models for diseases like Polycystic Ovary Syndrome (PCOS) using advanced computational techniques holds profound implications for society. By enabling early detection and personalized management strategies, these models can significantly improve healthcare outcomes for millions of individuals affected by PCOS worldwide. Timely identification of PCOS risk factors can lead to early interventions, reducing the burden of long-term complications such as infertility, diabetes, and cardiovascular diseases. Moreover, by leveraging machine learning algorithms, healthcare providers can deliver more tailored and effective treatment plans, ultimately enhancing the quality of life for patients while reducing healthcare costs and improving overall public health.

5.2 Impact on Environment

While the impact of predictive disease models on the environment may not be direct, their implementation can indirectly contribute to environmental sustainability through several avenues. By enabling early detection and personalized management of diseases like Polycystic Ovary Syndrome (PCOS), these models can lead to more efficient use of healthcare resources, reducing unnecessary medical procedures, prescriptions, and hospital admissions. This streamlined healthcare approach can minimize the environmental footprint associated with healthcare activities, including energy consumption, waste generation, and carbon emissions. Additionally, improved health outcomes and reduced disease burden can lead to a healthier population, potentially decreasing the overall demand for healthcare services and resources in the long term. Overall, while the environmental impact may be indirect, the implementation of predictive disease models has the potential to contribute to a more sustainable healthcare system and promote environmental stewardship.

5.3 Ethical Aspects

The development and deployment of predictive disease models, including those for conditions like Polycystic Ovary Syndrome (PCOS), raise several ethical considerations that must be carefully addressed. One primary concern is privacy and data protection, as these models often rely on sensitive personal health information. Ensuring the privacy and confidentiality of patient data through robust security measures and compliance with privacy regulations is essential to maintain trust and integrity in the healthcare system. Another ethical consideration is the potential for bias and discrimination in algorithmic decision-making. Predictive models may inadvertently perpetuate or exacerbate existing disparities in healthcare access and outcomes if not carefully designed and validated across diverse populations. It is crucial to mitigate bias through transparent model development, rigorous validation, and ongoing monitoring for fairness and equity. Furthermore, there are ethical implications surrounding informed consent and autonomy. Patients should be adequately informed about the purpose, risks, and limitations of predictive disease models and have the autonomy to consent or decline participation in such programs. Clear communication and patient education are essential to ensure informed decision-making and respect for individual preferences. Additionally, the responsible use of predictive disease models requires careful consideration of their impact on clinical practice and patient care. While these models can provide valuable decision support to healthcare providers, they should not replace human judgment or diminish the importance of patient-provider relationships. Ethical guidelines and standards of practice should be established to govern the integration of predictive models into clinical workflows and ensure their responsible use in patient care. Overall, addressing ethical considerations in the development and deployment of predictive disease models is essential to uphold principles of beneficence, non-maleficence, autonomy, and justice in healthcare delivery.

5.4 Sustainability Plan

The sustainability plan for predictive disease models targeting conditions like Polycystic Ovary Syndrome (PCOS) entails a multifaceted approach encompassing continuous improvement through research and development, adherence to ethical guidelines,

interdisciplinary collaboration, public education initiatives, capacity building in underserved communities, advocacy for supportive policies, ongoing evaluation and monitoring, and securing sustainable funding streams. By integrating these elements, stakeholders can ensure the long-term viability, ethical integrity, and societal benefit of predictive disease models while minimizing risks and promoting equitable access to healthcare innovations.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

The analysis of machine learning, bagging, boosting, stacking, voting, and deep learning models reveals distinct performance characteristics across various evaluation metrics. Machine learning models, including decision trees and logistic regression, demonstrate moderate performance, while ensemble methods such as bagging, boosting, and stacking showcase improvements in predictive accuracy and robustness. Bagging algorithms, such as Random Forest, enhance model stability by aggregating predictions from multiple decision trees, resulting in reduced variance and improved generalization. Boosting techniques, exemplified by Gradient Boosting, iteratively refine model predictions by emphasizing misclassified instances, leading to enhanced performance and predictive power. Stacking, a meta-ensemble method, combines the strengths of multiple base models through a higher-level learner, offering superior predictive performance compared to individual models. Voting, another ensemble approach, aggregates predictions from diverse models through a majority voting mechanism, resulting in improved accuracy and reliability. Deep learning models, including Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN), exhibit remarkable performance in capturing complex patterns and relationships within the data, leading to accurate predictions. Overall, the analysis underscores the versatility and effectiveness of ensemble methods and deep learning techniques in addressing diverse predictive modeling tasks, offering valuable insights for real-world applications.

6.2 Conclusions

In conclusion, the development and implementation of predictive disease models represent a significant advancement in healthcare, offering the potential to revolutionize early detection, personalized treatment, and overall patient care. Through the integration of

advanced computational techniques, ethical considerations, interdisciplinary collaboration, and ongoing evaluation, these models hold promise for improving health outcomes, reducing healthcare disparities, and enhancing the efficiency and sustainability of healthcare systems. However, to realize these benefits fully, it is essential to address challenges such as data privacy, bias mitigation, stakeholder engagement, and resource allocation, while fostering a culture of transparency, accountability, and continuous improvement. With concerted efforts from stakeholders across sectors, predictive disease models can play a pivotal role in shaping the future of healthcare, delivering more precise, equitable, and patient-centered services for the benefit of society as a whole.

6.3 Implication for Further Study

The findings of this study underscore the potential of predictive disease models, particularly in the context of Polycystic Ovary Syndrome (PCOS), to improve healthcare outcomes and patient care. However, several avenues for further research and exploration emerge from this study. Firstly, investigating the impact of different data preprocessing techniques and feature engineering methods on model performance could provide insights into enhancing predictive accuracy and generalization across diverse populations. Additionally, exploring novel machine learning architectures, such as attention mechanisms or graph neural networks, may further improve the interpretability and robustness of predictive models for PCOS. Furthermore, longitudinal studies evaluating the long-term effectiveness and sustainability of predictive models in clinical practice are warranted to assess their real-world utility and scalability. Moreover, research focusing on the integration of patient-reported outcomes, genetic data, and environmental factors into predictive models could advance our understanding of PCOS etiology and enable more personalized interventions. Lastly, interdisciplinary studies examining the societal, ethical, and economic implications of implementing predictive disease models in healthcare settings are crucial for guiding policy decisions and ensuring equitable access to these technologies. Overall, further research in these areas holds the potential to enhance the effectiveness, accessibility, and ethical integrity of predictive disease modeling for PCOS and other complex health conditions.

Reference:

[1] D. Hu, W. Dong, X. Lu, H. Duan, K. He and Z. Huang, "Evidential mace prediction of acute coronary syndrome using electronic health records", BMC Med. Informat. Decis. Making, vol. 19, no. 2, pp. 9-17, 2019.

[2] M. M. Hassan and T. Mirza, "Comparative analysis of machine learning algorithms in diagnosis of polycystic ovarian syndrome", Int. J. Comput. Appl., vol. 175, pp. 42-53, Sep. 2020.

[3] G. Du, L. Ma, J.-S. Hu, J. Zhang, Y. Xiang, D. Shao, et al., "Prediction of 30-day readmission: An improved gradient boosting decision tree approach", J. Med. Imag. Health Informat., vol. 9, no. 3, pp. 620-627, 2019.

[4] S. Bharati, P. Podder and M. R. H. Mondal, "Diagnosis of polycystic ovary syndrome using machine learning algorithms", Proc. IEEE Region Symp. (TENSYMP), pp. 1486-1489, Jun. 2020.

[5] Bhat, S. A. (2021). Detection of polycystic ovary syndrome using machine learning algorithms (Doctoral dissertation, Dublin, National College of Ireland).

[6] S. Yang, X. Zhu, L. Zhang, L. Wang and X. Wang, "Classification and prediction of Tibetan medical syndrome based on the improved bp neural network", IEEE Access, vol. 8, pp. 31114-31125, 2020.

[7] D. Dewailly, M. E. Lujan, E. Carmina, M. I. Cedars, J. Laven, R. J. Norman, et al., "Definition and significance of polycystic ovarian morphology: A task force report from the androgen excess and polycystic ovary syndrome society", Hum. Reproduction Update, vol. 20, no. 3, pp. 334-352, 2014

[8] A. Saravanan and S. Sathiamoorthy, "Detection of polycystic ovarian syndrome: A literature survey", Asian J. Eng. Appl. Technol., vol. 7, pp. 46-51, Nov. 2018

[9] X.-Z. Zhang, Y.-L. Pang, X. Wang and Y.-H. Li, "Computational characterization and identification of human polycystic ovary syndrome genes", Sci. Rep., vol. 8, no. 1, Dec. 2018.

[10] V. Thakre, "PCOcare: PCOS detection and prediction using machine learning algorithms", Biosci. Biotechnol. Res. Commun., vol. 13, no. 14, pp. 240-244, Dec. 2020.

[11] R. M. Aziz, "Nature-inspired metaheuristics model for gene selection and classification of biomedical microarray data", Med. Biol. Eng. Comput., vol. 60, no. 6, pp. 1627-1646, 2022.

[12] R. M. Aziz, "Application of nature inspired soft computing techniques for gene selection: A novel frame work for classification of cancer", Soft Comput., vol. 1, pp. 1-18, Apr. 2022.

[13] Z. Na, W. Guo, J. Song, D. Feng, Y. Fang and D. Li, "Identification of novel candidate biomarkers and immune infiltration in polycystic ovary syndrome", J. Ovarian Res., vol. 15, no. 1, pp. 1-13, 2022.

[14] S. Dhar, S. Mridha and P. Bhattacharjee, "Mutational landscape screening through comprehensive in silico analysis for polycystic ovarian syndrome–related genes", Reproductive Sci., vol. 29, no. 2, pp. 480-496, 2022.

[15] Shorove Tajmen, Asif Karim, Aunik Hasan Mridul, Sami Azam, Pronab Ghosh, Alamin Dhaly, Md Nour Hossain. “A Machine Learning based Proposition for Automated and Methodical Prediction of Liver Disease”. In April 2022 The 10th International Conference on Computer and Communications Management in Japan.

[16] Aunik Hasan Mridul, Md. Jahidul Islam, Mushfiqur Rahman, Mohammad Jahangir Alam, Asifuzzaman Asif. “A Machine Learning-Based Traditional and Ensemble Technique for Predicting Breast Cancer”, In December, 2022. Conference: 22th International Conference on Hybrid Intelligent Systems (HIS 2022) online, 2022At: Auburn, Washington, USA.

[17] “Logistic Regression for Machine Learning”, Accessed: August 6, 2021, Available: https://www.capitalone.com /tech/machine-learning/what-is-logistic-regression/

[18] Ghosh, Pronab, Asif Karim, Syeda Tanjila Atik, Saima Afrin, and Mohd Saifuzzaman. "Expert cancer model using supervised algorithms with a LASSO selection approach." International Journal of Electrical and Computer Engineering (IJECE) 11, no. 3 (2021): 2631

[19] emmens, Aurélie, and Christophe Croux. "Bagging and boosting classification trees to predict churn." Journal of Marketing Research43, no. 2 (2006): 276-286.

[20] Bentéjac, Candice, Anna Csörgő, and Gonzalo Martínez-Muñoz. "A comparative analysis of gradient boosting algorithms." ArtificialIntelligence Review 54, no. 3 (2021): 1937-1967.

[21] emmens, Aurélie, and Christophe Croux. "Bagging and boosting classification trees to predict churn." Journal of Marketing Research43, no. 2 (2006): 276-286

[22] Wang, Yizhen, Somesh Jha, and Kamalika Chaudhuri. "Analyzing the robustness of nearest neighbors to adversarial examples." In International Conference on Machine Learning, pp. 51335142. PMLR, 2018
[23] Drucker, Harris, Corinna Cortes, Lawrence D. Jackel, Yann LeCun, and Vladimir Vapnik. "Boosting and other ensemble methods." Neural Computation 6, no. 6 (1994): 1289-1301
[24] Sharma, Ajay, and Anil Suryawanshi. "A novel method for detecting spam email using KNN classification with spearman correlation as distance measure." International Journal of Computer Applications 136, no. 6 (2016): 28-35
[25] Pasha, Maruf, and Meherwar Fatima. "Comparative Analysis of Meta Learning Algorithms for Liver Disease Detection." J. Softw. 12, no.12 (2017): 923-933
[26] Islam, Rakibul, Abhijit Reddy Beeravolu, Md Al Habib Islam, Asif Karim, Sami Azam, and Sanzida Akter Mukti. "A Performance Based Study on Deep Learning Algorithms in the Efficient Prediction of Heart Disease." In 2021 2nd International Informatics and Software Engineering Conference (IISEC), pp. 1-6. IEEE, 2021
[27] Aunik Hasan Mridul, Nowreen Ahsan, Md. Abu Saleh, Asifuzzaman, Sonia Afrose. “Effective Early Hypothyroid Disease Prediction Using Traditional and Ensemble Machine Learning Algorithms”, SoCPaR 2023, December 2023.
[28] P. Kottarathil, "Polycystic ovary syndrome (PCOS) |Kaggle", 2022

DISEASE PREDICTION FOR POLYCYSTIC OVARY SYNDROME (PCOS) USING DEEP LEARNING AND CONVENTIONAL MACHINE LEARNING ALGORITHMS

ORIGINALITY REPORT

21%	15%	13%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	3%
2	Shazia Nasim, Mubarak Saad Almutairi, Kashif Munir, Ali Raza, Faizan Younas. "A Novel Approach for Polycystic Ovary Syndrome Prediction Using Machine Learning in Bioinformatics", IEEE Access, 2022 Publication	2%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
4	thesai.org Internet Source	2%
5	Shazia Nasim, Mubarak Almutairi, Kashif Munir, Ali Raza, Faizan Younas. "A Novel Approach for Polycystic Ovary Syndrome Prediction Using Machine Learning in Bioinformatics", IEEE Access, 2022 Publication	1%