PREDICTIVE-PROBABILITY ANALYSIS AND PREVENTIVE STRATEGIES FOR HEMORRHOIDS: A MACHINE LEARNING APPROACH

BY

Sajib Saha
ID: 202-15-3791
AND

Tahsin Jahan
ID: 202-15-3804

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised By

Israt Jahan
Lecturer (Senior Scale)
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Ms. Fabliha Haque
Lecturer
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled **"Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach"**, submitted by **Sajib Saha** and **Tahsin Jahan** 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 **13-07-2024**.

BOARD OF EXAMINERS

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

13.07.24

Ms. Nazmun Nessa Moon — **Internal Examiner 1**
Associate Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

13.7.24

Mr. Deawan Rakin Ahamed Remal — **Internal Examiner 2**
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

13.07.2024

Dr. Abu Sayed Md. Mostafizur Rahaman — **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Israt Jahan, Lecturer (Senior Scale), Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

11.7.24

Israt Jahan
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Co-Supervised by:

11.7.24

Ms. Fabliha Haque
Lecturer
Department of CSE
Daffodil International University

Submitted by:

SajibSaha

Sajib Saha
ID: 202-15-3791
Department of CSE
Daffodil International University

Tahsin Jahan

Tahsin Jahan
ID: 202-15-3804
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

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

We are grateful and wish our profound indebtedness to **Israt Jahan**, **Lecturer (Senior Scale)**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach” to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

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

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

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

In the world millions of people suffer from hemorrhoids. It is a medical condition that frequently requires prompt intervention for effective management. In this paper, we suggest a machine learning-based method for estimating an individual's risk of developing hemorrhoids. We are using a dataset that includes lifestyle factors, medical history, and demographic data. We use five algorithms: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest, XGBoost and Naive Bayes to create prediction models. Cross-validation and feature selection strategies are applied to improve the validity and performance of the model in this paper through dataset. Our research shows encouraging outcomes in precisely identifying people who are susceptible to hemorrhoids, which will enable early interposition and individualized treatment plans. This work offers a proactive approach to hemorrhoid management and its prevention system, furthering the field of predictive analysis in healthcare. Random Forest's accuracy 95.19%, KNN's accuracy is 95.21%, SVM's accuracy is 95.93%, XGBoosting's accuracy is 95.32% and Naive Bayes's accuracy is 90.37. SVM's training accuracy is the highest and it is 96.67% So, SVM is suitable and best algorithm for predictive-probability analysis and preventive strategies for hemorrhoids. We have 1350 primary data. The dataset consists of 670 of patient records and 680 of normal records. Additionally, a brief summary of preventive strategies intended to lessen the frequency and severity of hemorrhoids is included in our research paper. This paper seeks to equip people with the knowledge and skills needed to reduce the burden of hemorrhoids and improve their overall quality of life by spreading awareness about preventive measures.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Overview

Hemorrhoids, a common medical condition characterized by swollen blood vessels in the rectal area, affect millions of individuals globally, leading to significant discomfort and impacting quality of life. Traditional methods of diagnosing and predicting hemorrhoids rely heavily on symptomatic assessment and invasive procedures, which can be time-consuming and often result in delayed intervention. With advancements in technology, machine learning offers a promising alternative by providing accurate, non-invasive prediction models based on vast datasets comprising patient demographics, medical histories, lifestyle factors, and other relevant parameters. This research paper explores the development and application of machine learning algorithms to predict the likelihood of hemorrhoids, aiming to enhance early detection and personalized treatment plans. By integrating these predictive models into clinical practice, healthcare providers can improve patient outcomes, optimize resource allocation, and reduce the overall burden of this prevalent condition. This study underscores the transformative potential of machine learning in medical diagnostics, particularly in addressing the challenges associated with hemorrhoids.

1.2 Background and Present State

In order to simplify a clear understanding of the key ideas and methodologies employed in this paper on "Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach". It is dominant to establish a foundation through introductions and the illumination of terms. It is essential to establish an initial understanding of key concepts. Firstly, hemorrhoids, often mentioned to as piles, are explained as swollen veins surrounding the rectum or anus, leading to distress, bleeding, and other symptoms. Next, symptoms and danger issues related with hemorrhoids are outlined, highlighting the importance of classifying individuals at risk. The section proceeds to introduce central machine learning ideas, including predictive modeling, feature selection, and evaluation metrics. Understanding these machine learning values is vital for developing accurate prediction models. Moreover, the concept of cross-validation is introduced as a method for measuring model act and confirming generalizability. By clarifying these terms, the basis is laid for discovering the application of machine learning in predicting hemorrhoids an important role to healthcare and defensive remedy. A thorough analysis of related works and earlier research is provided in the background chapter. A general grasp of the abstract keystones, technical developments, and methods associated with hemorrhoid detection and machine learning is provided in this section. It provides background information for the current study and finds information gaps.

1.3 Problem Statement

Nowadays Hemorrhoids is a very common disease. Recently not only old people but also young people are getting affected by this disease. Nowadays most of the young people eat heavy food like fish, meat, fry and many more. People now also drink less water and eat less vegetables. These are the some causes of hemorrhoids. Robust predictive tools are currently lacking in clinical practices to identify patients who are more susceptible to hemorrhoids. Reactive rather than proactive management results from the lack of a systematic approach to identify predisposing factors and patterns, placing a significant burden on patients and healthcare systems. Furthermore, developing a reliable predictive framework is difficult due to the multifactorial nature of hemorrhoids, which includes genetic predisposition, lifestyle factors, and anatomical variations. Current research emphasizes the urgent need for a comprehensive predictive model to enable early detection and personalized preventive strategies, as it focuses primarily on treatment rather than prediction. It is critical to close this gap in hemorrhoid predictive analysis in order to transform healthcare delivery, allocate resources optimally, and greatly enhance patient outcomes. Thus, the goal of this research is to create a reliable prediction model that integrates clinical, lifestyle, and demographic data in order to proactively identify people who are at risk of hemorrhoids. This will enable medical professionals to take preventative action and enhance patient outcomes.

1.4 Objectives

A significant percentage of the population suffers from hemorrhoids, a common ailment that frequently causes discomfort, lowers quality of life, and, in extreme situations, complications that call for medical attention. The urgent need to create a trustworthy prediction model is highlighted by the dearth of preventative measures and predictive tools in clinical practice. By enabling early identification of high-risk individuals, facilitating timely interventions, and optimizing healthcare resources, this model seeks to transform the way that hemorrhoids are treated. Using machine learning for the identification, diagnosis, and management of hemorrhoids is an emerging area of interest in medical research. Here are some examples of possible goals for our paper:

- Classification and Diagnosis Improvement.
- Predictive Models for Symptom Severity.

- Data Analysis and Pattern Recognition.
- Treatment Outcome Prediction.
- Risk Stratification and Prevention Strategies.
- Patient-Centric Decision Support Systems.
- Data-driven Insights for Healthcare Policies.
- User-friendly Tools for Self-Assessment and Management.
- Improving Public Health Awareness.
- Enhancing Preventive Healthcare Strategies

Each of these goals seeks to maximize the potential of machine learning in order to improve treatment outcomes, optimize preventive measures, increase diagnostic accuracy, or offer more insights into the management of hemorrhoid diseases. The overall goal of this research is to use machine learning for predictive analysis in the field of healthcare, with a particular focus on hemorrhoids, and to extract useful information that may be used to prevent or lessen the effects of this problem.

1.5 Scope and Limitations

This research aims to develop and validate machine learning models for predicting the occurrence of hemorrhoids, leveraging comprehensive datasets that include patient demographics, medical histories, lifestyle factors, and other pertinent variables. The primary objective is to create a predictive tool that can be seamlessly integrated into clinical workflows, thereby aiding healthcare providers in early diagnosis and personalized treatment planning. The scope of the study includes the selection of appropriate machine learning algorithms, preprocessing of data, model training and validation, and assessment of model performance against established benchmarks.

However, several limitations must be acknowledged. The accuracy and reliability of the predictive models are intrinsically linked to the quality and completeness of the data used. Incomplete, biased, or unrepresentative datasets can lead to flawed predictions and reduced generalizability. Additionally, the applicability of these models may vary across different patient populations and healthcare settings, potentially limiting their effectiveness in diverse clinical environments. Ethical considerations, particularly concerning patient privacy and data security, are critical challenges that must be addressed to maintain trust and compliance with regulatory standards. Furthermore, the successful implementation of these models in real-world settings requires extensive validation, acceptance, and adoption by the medical community, which can be a time-consuming and complex process. Despite these challenges, the potential of machine learning to transform hemorrhoids prediction and improve patient outcomes highlights the importance of continued research and development in this area.

1.6 Report Organization

Through the procedure, conclusions, and results of the study, the suggested report has a thorough format, walks readers. Every chapter has a separate function and adds to the documents overall unity and depth.

Chapter 1: Introduction

The study is set in motion by the introductory chapter which gives background information on the importance of hemorrhoid detection, the requirement for a predictive-probability analysis of detection techniques and the use of machine learning. It provides an overview of the paper's areas, research questions, and general framework.

Chapter 2: Literature Review

A thorough analysis of related works and earlier research is provided in the background chapter. A general grasp of the abstract keystones, technical developments, and methods associated with hemorrhoid detection and machine learning is provided in this section. It provides background information for the current study and finds information gaps.

Chapter 3: Methodology/ Requirement Analysis & Design Specification

The approach used to carry out the study is described in the research methodology chapter. It sheds light on the model planning, data collection systems, study design, and the reasoning behind particular methodology collections. The moral issues and potential limits on the study are also enclosed in this section.

Chapter 4: Implementation

The implementation section details the construction of our machine learning model for hemorrhoid prediction. We evaluated five algorithms: Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), XGBoost and Naive Bayes. The data underwent preprocessing to ensure its quality for analysis, including techniques for handling missing values and potentially creating new informative features. The chosen algorithms were then trained on the prepared data, essentially allowing them to learn patterns associated with hemorrhoids. Finally, we employed evaluation metrics to assess each model's performance in predicting hemorrhoid presence.

Chapter 5: Result and Analysis

The paper begins by brief the performance metrics, including accuracy, precision, memory, F1-score, and area under the receiver operating characteristic curve for each machine learning algorithm considered. Relative analysis is then conducted to compare the performance of different algorithms and classify the most effective approach for hemorrhoid prediction. Overall, the new results and analysis provide valuable visions into the efficiency of machine learning models for hemorrhoid prediction using the primary dataset, controlling future research directions and informing medical decision-making methods.

Chapter 6: Impact on Society, Environment and Sustainability

The allegations of the research on "Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach " range far beyond the empire of academia, promising significant impacts on society and healthcare. Predictive models can help healthcare providers identify individuals at risk of developing hemorrhoids earlier, allowing timely interpolation and dealing. This can lead to improved healthcare access and results for patients. Early detection and interpolation can possibly decrease the need for costly medical actions related with advanced hemorrhoids.

Chapter 7: Conclusion and Further Work

In conclusion, this paper reveals the possible of machine learning techniques in predicting the rate of hemorrhoids. By studying various clinical and lifestyle issues, we have developed predictive models with capable accuracy rates. These models not only help in early identification of persons at risk of emerging hemorrhoids but also deliver valuable visions for preventive healthcare strategies.

1.7 Summary

In summary, this introduction lays the foundation for exploring the application of machine learning in predicting hemorrhoids, highlighting its potential to revolutionize healthcare by facilitating early detection and personalized treatment strategies. The significance of this research lies in its potential to alleviate the burden of hemorrhoids through non-invasive predictive models, thereby improving patient outcomes and optimizing healthcare resources. The scope includes the development and validation of machine learning algorithms using diverse datasets, while acknowledging the limitations such as data quality, model generalizability, and ethical considerations. This paper proceeds to review relevant literature, detail the methodology employed, present and analyze findings, discuss implications for clinical practice, and suggest avenues for future research. By addressing these aspects, this study aims to contribute valuable insights into leveraging advanced technology for enhancing medical diagnostics and patient care in the context of hemorrhoids management.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

The literature review section provides a comprehensive overview of existing research and methodologies related to the prediction and management of hemorrhoids, focusing particularly on the application of machine learning techniques. It examines studies investigating traditional diagnostic methods, their limitations, and the growing interest in leveraging data-driven approaches for early detection and intervention. Key topics include the epidemiology of hemorrhoids, clinical manifestations, and current medical practices in diagnosis and treatment. The review also explores recent advancements in machine learning algorithms, such as supervised learning models logistic regression, decision trees, support vector machines and deep learning techniques neural networks, highlighting their potential to enhance predictive accuracy and clinical decision-making. Furthermore, this section discusses the challenges and gaps in existing literature, setting the stage for the current study's contribution in bridging these gaps through innovative application of machine learning in hemorrhoids prediction.

2.2 Related Works

This item discusses the basic understanding and current management of hemorrhoids from the viewpoint of a proctologist. This contains managing hemorrhoids in patients with cirrhosis or portal hypertension, hemorrhoids in patients receiving antithrombotic agents, hemorrhoids in pregnant women, hemorrhoids in immunocompromised patients, and hemorrhoids that are acutely thrombosed or suffocated. Future directions in hemorrhoid treatment are also covered [1].

Hemorrhoids are lumps or clumps of tissue that are found in a person's anal canal. They are also known as piles. They are made of elastic and muscle fibers, with surrounding supportive tissues and enlarged, projecting blood vessels. This is a condition that is common in adulthood. The disease's true burden is yet unknown. Therefore, the author conducted this study to evaluate the disease's incidence and risk factors among local patients. The study comprised sixty-three patients with hemorrhoids who visited the outpatient ward between the ages of 20 and 80. Every patient's demographic information was recorded. Their consumption of alcohol and cigarettes, their bowel movements, their level of physical activity, and their usage of over-the-counter medications were all meticulously documented [2].

It is rare to find precise statistics regarding the incidence of hemorrhoids. In order to investigate the prevalence of hemorrhoids and the risk factors related to them in the adult general population, this publication suggests conducting a study. Of the 976 participants in the study, 380 (38.93%) had hemorrhoids. In 277 patients (72.89%), grade I hemorrhoids were diagnosed; in 70 patients (18.42%), grade III in 31 patients (8.16%), and grade IV in 2 patients (0.53%), hemorrhoids were diagnosed as grade I. Among the patients, 210 (55.26%) reported no symptoms, and 170 (44.74%) reported symptoms related to hemorrhoids. In both the univariate and multivariate analyses, body mass index (BMI) significantly influenced the incidence of hemorrhoids ($p = 0.0391$ and $p = 0.0282$, respectively). Even after controlling for other potential risk factors, a one-point rise in BMI raised the chance of hemorrhoids by 3.5%. Furthermore, a high BMI can be considered an independent risk factor for hemorrhoids [3].

Based on the most current research and reviews of the literature, this article addresses contemporary methods for treating hemorrhoids. The current pathologies associated with hemorrhoids include vascular hyperplasia, hemorrhoid plexus hyperperfusion, and degenerative changes in the supportive tissue inside the anal cushions. A combination of medical intervention, dietary and lifestyle modifications, and office-based procedures can effectively treat low-grade hemorrhoids in a timely manner. An operation is often advised in the case of symptomatic high-grade and/or complicated hemorrhoids. The treatment of hemorrhoids in difficult cases is also covered [4].

The main finding of this study was the percentage of hemorrhoids at the time of discharge from the obstetrics unit. To compare categorical variables, the Fisher exact test or the Chi-Squared test were utilized. Continuous variables were compared using the Mann-Whitney U test or the Student's t-test. The study utilized a binary logistic regression model to determine the independent factors associated with postpartum hemorrhoids. This analysis was conducted on factors with a p-value of 0.10 in univariate analysis. For statistical analysis, GraphPad Prism 9 and IBM SPSS 23.0 were utilized. A P-value of less than 0.05 was deemed significant for all tests [5].

Randomized prospective trials and meta-analyses were conducted in this research to determine the gold standard for surgery. When used to treat advanced disease, HC appears to be more effective than PPH, which produces decent results in third-degree piles but has a high reintervention rate in fourth-degree piles. Almost all experiments comparing open and closed HC have comparable results. Both HC and PPH are unlikely to result in permanent or substantial anal incontinence. Anal dilatation can relieve the majority of cases of anal stricture caused by HC. According to society's standards, patients should have customized surgery, meaning that different techniques should be used depending on the severity of hemorrhoids. This means that patients should be operated on by a professional colorectal surgeon who is qualified to do various surgeries and handle complications and failures [6].

The current study looks at changes in the caliber and flow of the terminal branches of the SRA, which supply the corpus cavernosum recti (CCR), in hemorrhoid patients. Transperineal color Doppler ultrasound was used to compare 41 outpatients (17 female, 24 male; mean age 48 years) with Goligher grades I–IV hemorrhoids with 17 healthy volunteers (9 female, 8 male; mean age 29 years). The study group with hemorrhoids had an average arterial branch caliber of 1.87 6 0.68 mm (range, 0.6 to 3.60 mm), while the control group had an average caliber of 0.92 6 0.15 mm (range, 0.6 to 1.2 mm) ($P < 0.001$). Their results show a association between the occurrence of hemorrhoids and increased major blood flow and ability of the SRA terminal twigs. Rather than being a result of hemorrhoids, they propose that the hyper vascularization of the anorectum is a contributing factor to the development of hemorrhoids. In patients with hemorrhoids, transperineal color. Doppler ultrasonography (CDUS) is a appropriate technique to assess these results [7].

The determination of the current study was to determine the occurrence of complaints among patients with hemorrhoids and to assess any potential accompaniment between hemorrhoidal disease and chronic venous disease. Small-scale data on patient features, risk factors, and often suggested treatments for people with hemorrhoids is provided by the paper. The meaning of checking for chronic venous disease (CVD) in conjunction with a hemorrhoidal analysis is underlined by the 13.8% of hemorrhoidal patients who also had chronic venous disease. Large-scale data and additional clinical research are required to confirm the co-morbidity of these two illnesses, and as a result, a shared targeted treatment should be created. A cautious approach seems to be the cornerstone of managing hemorrhoidal patients' lives, as even the earliest symptoms have a significant impact [8].

This research presents a methodology that predicts the bearing capacity of precast concrete piles using machine learning models as an alternative to traditional techniques. It makes use of 165 load test data samples, each of which has an SPT sound. Using two distinct datasets, this study suggests two different ways of analysis: one based on the Décourt-Quaresma approach and the other on the Meyerhof method. A leave-one-out cross-validation technique is used to train and test six machine learning algorithms with different biases. The predictive performance of the models is evaluated using the coefficient of determination (R2) and root mean squared error (RMSE). With an RMSE of 642.38, random forest (RF) utilizing the Décourt-Quaresma dataset proved to be the most effective method. The semi-empirical approaches of Meyerhof and Décourt-Quaresma, which both produced RMSE values close to 900, were defeated by all other machine learning techniques, which all produced an RMSE below 710. This study suggests using a graphical interpretation to provide 95% and 90% confidence intervals for the optimal technique, allowing geotechnical engineers to select the level of safety that best suits their needs. The paper concludes with a case study that demonstrates how, in two of the three piles examined, the best-performing models outperform the semi-empirical methods in terms of accuracy. The accuracy of the strategies is explained by the representatively of the new examples within the datasets used [9].

The goal of this research was to conduct an objective assessment of this novel technology by utilizing an ultrasonically triggered scalpel as a substitute for closed hemorrhoidectomy. In patients undergoing Harmonic Scalpel, the mean operative time was 31.7 + 2 minutes ($P < 0.37$); for closed hemorrhoidectomy with electrocautery, it was 35.7 + 3 minutes. The operating time for a two- or three-column hemorrhoidectomy did not differ statistically. While the Harmonic Scalpel is a useful tool for treating hemorrhoidal illness, when compared to a typical closed hemorrhoidectomy, there was no discernible improvement in terms of postoperative discomfort, faucal incontinence, operating time, quality of life, or complications [10].

This article goes over a few of these endoscopic methods, including rubber band ligation, sclerotherapy, and electrocoagulation. These treatments' development, efficacy, and benefits are summarized and reviewed. Endoscopic technology is projected to be more widely used in clinical practice in the future, possibly becoming the first-line therapy for hemorrhoids [11].

This is a prospective study of 50 patients with prolapsed piles, 46 of whom were males and 4 were girls, gathered over a two-year period from Sulaymaniyah Teaching Hospital and Shar Hospital in Iraq's Kurdistan Region. Mannitol was applied topically to the prolapsed hemorrhoids using gauze soaked in mannitol solution. All patients had prolapsed hemorrhoids, 46 of whom were males and 4 of whom were females. For data analysis, "IBM SPSS Statistics version 25" was utilized, along with descriptive and inferential statistics. Furthermore, P-values of 0.05 and 0.001 were regarded as statistically significant and highly significant, respectively. Furthermore, the Pearson Chi-Square Test was employed to establish whether there was a significant relationship between independent and dependent variable pairs. Prolapsed hemorrhoids can be handled conservatively by using mannitol topically which reduces edema and causes the hemorrhoidal tissue to retract to its original position. It is a simple, safe, and successful alternative to surgical therapy for prolapsed hemorrhoids [12].

These texts explained symptomatic hemorrhoids as the consequence of faces infecting the veins in the rectum, raising the vein's temperature, and causing it to enlarge. Using a red-hot iron to cauterize the hemorrhoids could result in a successful outcome. At the Battle of Waterloo in 1815, the British ultimately overthrew Napoleon. Napoleon's hemorrhoids proved to be a decisive factor in the fight, according to multiple stories from individuals who knew him well. Unlike how he usually conducts himself in combat, he rested mostly not on his horse but rather on a mountaintop, watching the action. He struggled when he tried to walk. Moreover, he had been using three to four leeches on a regular basis to treat his hemorrhoids, according to a letter he wrote to his brother a few years prior to the Battle of Waterloo. Nowadays, there are over 65,000 websites to choose from while researching hemorrhoids on the Internet [13].

They searched the literature for all publications written in English that dealt with hemorrhoids and were released between 1990 and 2002. MEDLINE, PreMEDLINE, the Cochrane Central Registry of Controlled Trials, the American College of Physicians Journal Club, the Database of Abstracts of Reviews of Effectiveness, and the Cochrane Database of Systematic Reviews were among the databases that were searched. The bibliographies of a few chosen articles provided more references. With an emphasis on randomized controlled trials, we selected relevant studies to create this technical review [14].

Evaluation of the countless treatment options for hemorrhoids is unfeasible given the lack of an internationally recognized classification system. A new classification system is put out that links symptoms to what is known about hemorrhoidal development. This system should make it easier to compare treatments and assess novel approaches. A thorough presentation of the evidence supports this classification [15].

A description is given of an anatomical and clinical investigation that sought to identify variables that would be useful in determining the true nature of hemorrhoids. It's said that piles are only the consequence of the specialized "cushions" of submucosal tissue bordering the anal canal, which constitute the principal finding. Though presently accepted, there are several theories that explain hemorrhoids other than varicosity of the anal submucosal veins, and this idea has not always been the most widely accepted. In fact, throughout the previous century, the majority of European authorities believed that piles were produced by vascular tissue hyperplasia, a theory that appears to be supported by current research conducted in Germany. According to a different theory, piles are just the result of the anal canal lining moving lower [16].

Cryotherapy and a host of additional devices intended to eliminate hemorrhoidal tissue with minimal postoperative pain are on this list. These days, bipolar scissors, diathermy instruments, carbon dioxide and neodymium yage lasers, and, more recently, the monopole coagulator (LigaSure, Valleylab, Boulder, CO) and the ultrasonic scalpel are used instead of the traditional Milligan Morgan or Ferguson hemorrhoidectomy. All of these approaches have one thing in common: despite early enthusiasm, thorough scientific examination was unable to validate the preliminary findings, which prevented broad adoption. Examples include Smith and colleagues' cryotherapy experiment as well as Leff's, Senagore's, and associates' experiences with laser hemorrhoid. The latter study was, in fact, a well designed, prospectively randomized experiment where longer wound healing in the laser group was the sole noteworthy distinction between the usual excisional approach and the laser [17].

The investigation of concomitant hemorrhoids and other illnesses in identical people may point to a common underlying etiology. The current investigation sought to ascertain which diagnoses are linked to the development of hemorrhoids. Constipation does not appear to be more common in patients with hemorrhoids than in control subjects without hemorrhoids. In fact, hemorrhoids were reported to be associated with diarrhea rather than constipation. 6 The United States Department of Veterans Affairs (VA) keeps computerized records on all patients who have received care at any of the VA hospitals across the country. A huge database like this provides an once-in-a-lifetime opportunity to examine the comorbid incidence of various diseases. The current investigation sought to ascertain which diagnoses are linked to the development of hemorrhoids. [18].

Open hemorrhoidectomy complications were uncommon. Anal continence remained unchanged. One year after surgery, comfort and well-being had greatly improved. Despite the presence of lingering anal discomfort, patient satisfaction was good. Surgical hemorrhoidectomy has long been the conventional treatment for hemorrhoids, but there is little data on how it affects symptoms in individuals who are followed over time. The current investigation found that problems are uncommon. During the follow-up period, hemorrhoidal symptoms, continence, and continence disturbance improved. Despite the occasional persistence of hemorrhoidal symptoms in some cases, patient satisfaction was excellent [19].

This research aims to evaluate the efficacy of CEH in comparison to Ferguson hemorrhoidectomy (FH) for end-stage hemorrhoids. 688 patients who arrived with end-stage hemorrhoids and underwent FH or CEH at our center between February 1998 and December 2011 were included in a prospective randomized experiment. Outcomes The study groups were comparable in terms of patient demographics, lengths of hospital stay, mean operating times, and cumulative incidence of postoperative complications. There were notable variations in the incidence of anal stricture (15 vs. 32 patients in the FH and CEH groups, respectively; $p = 0.02$) and postoperative hemorrhage (9 vs. 0 patients in the FH and CEH groups, respectively; $p = 0.002$). All cases of anal strictures, on the other hand, were simply handled by digital dilatations. The CEH group remained symptom-free after a mean follow-up of 7.4 (range, 1–14) years, whereas 126 of the 308 patients in the FH group reported recurring hemorrhoidal symptoms. In terms of reducing the rate of recurrence to zero by total hemorrhoid ectopy, CEH outperforms FH without increasing postoperative complications [20].

2.3 Comparison Between Existing Works

The comparative analysis section for the "Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach" research paper provides a brief overview of existing methods and methods in predicting hemorrhoids. It associates traditional medical techniques with new progresses in machine learning. The section highlights the boundaries of current methods and stresses the need for more accurate critical models. By brief key findings from related literature, this section sets the stage for the research by representative the gap in current approaches and modifying the use of machine learning for hemorrhoids prediction. Also, the section discovers the latent benefits of leveraging machine learning algorithms for hemorrhoids prediction. It confers how machine learning techniques by examining large datasets and identifying complex patterns, can offer more accurate and dependable forecasts compared to traditional approaches. Furthermore, the section may summarize key findings from previous research papers or lessons that have inspected the application of machine learning in other medical realms related to abdominal diseases or predictive modeling. Provide a concise overview of the studies you identified through your search. This might involve summarizing the number of studies, the type of machine learning algorithms used, the features analyzed, and the reported performance metrics (accuracy, sensitivity, specificity). Compare the different machine learning algorithms used in the studies. Discuss the rationale behind these features and how they might be relevant to hemorrhoid development. Compare the performance metrics reported in the studies. Discuss how well the different algorithms performed and any potential reasons for variations in performance. Overall, this section serves to contextualize the research by providing a comprehensive overview of existing methodologies, thereby underscoring the necessity and potential advantages of employing machine learning for hemorrhoids prediction.

Table 2.3.1: Comparative Analysis Table

Author	Algorithm	Dataset	Result	Contribution	Limitation
Stefan Riss, Friedrich Anton Weiser[3]	Random Forest	Primary Dataset	72.89%	The study examined hemorrhoid prevalence and related risk factors in the adult population.	Studying adult patients might have resulted in more educated patients.
Tomas Poskus, Grazina Drasutiene[5]	Logistic Regression, KNN	Primary Dataset	76%	The study compared dietary and behavioral interventions to prevent hemorrhoids during pregnancy in a multicenter, randomized clinical trial.	Patients compliance with intervention recommendations was not strictly monitored.

Yago Ferreira Gomes1[9]	Random Forest, KNN	Document Dataset	Random Forest = 95%	Graphical presentation of 95% and 90% confidence intervals for the optimal technique, aiding geotechnical engineers in selecting desired safety levels.	The study only takes into account a small number of SPT soundings and load tests.
M.D. Khan S.Eggenberger Margolin[10]	Random Forest	Primary Dataset	65%	The paper compared ultrasonically activated scalpel to closed hemorrhoid, evaluating operative time, postoperative pain, incontinence, quality of life, and complications.	Study only assessed short term complications, not long term effects.
Wenzhuang Ma1 ,Fan Yang1[11]	Logistic Regression	Primary Dataset	64%	It focuses on flexible reversible endoscopes, enhancing treatment outcomes through precise control and imaging.	Hemorrhoids are not discussed in the paper in terms of their limitations or drawbacks.
Sheikh FirojKabir, Debasish Das[8]	KNN, Random Forest	Clinical Dataset	71%	The paper aims to assess complaint frequency in hemorrhoidal patients and explore the relationship between hemorrhoidal disease and chronic venous disease.	The study was conducted in multiple centers with different physicians, but the communication between the physicians was not strong.
Fabiola Delc[18]	Random Forest	Synthetic Dataset	73%	The paper aimed to identify diagnoses associated with hemorrhoid occurrence through a case-control study.	This paper does not discuss the specific limitations and biases of the research.

2.4 Open Issues

Current research on “Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach” faces several open issues. Data availability and quality remain significant challenges, as obtaining large, diverse, and high-quality datasets is difficult due to privacy concerns and variability in sources. Feature selection and data preprocessing require careful handling to ensure relevant and consistent inputs. Algorithm-specific challenges also arise: KNN struggles with computational inefficiency, SVM with kernel selection, Random Forest with interpretability XGBoosting and Naive Bayes with overfitting. Moreover, model evaluation lacks standardization, making comparisons difficult. Ensuring model interpretability and integration into clinical practice is crucial for clinician trust and usability. Ethical and legal considerations, including data privacy and bias, must be addressed. Ensuring model generalizability and robustness across different populations and settings is essential. Finally, continuous monitoring and updating of models are necessary to maintain their relevance and accuracy over time. Addressing these issues will enhance the development and application of machine learning models for hemorrhoid prediction, ultimately improving patient care.

Data Availability and Quality:

- Lack of large, diverse, and high-quality datasets specific to hemorrhoid prediction.
- Difficulty in acquiring comprehensive medical records due to privacy concerns and ethical considerations.
- Variability in data sources that can introduce inconsistencies.

Feature Selection and Data Preprocessing:

- Identification of relevant features that significantly impact hemorrhoid prediction.
- Challenges in handling missing, incomplete, or noisy data.
- Need for standardization in data preprocessing techniques across different studies.

Model Evaluation and Validation:

- Lack of standard evaluation metrics across different studies.
- Need for comprehensive validation strategies, including cross-validation and external validation.
- Difficulty in comparing results due to variations in study designs and datasets.

Ethical and Legal Considerations:

- Ensuring patient data privacy and security during model development and deployment.
- Addressing potential biases in models that could lead to unfair or inaccurate predictions.

- Compliance with legal regulations and standards in healthcare data usage.

Long-Term Model Maintenance:

- Establishing protocols for continuous monitoring and updating of models to maintain accuracy over time.
- Adapting models to evolving medical knowledge and changes in clinical practices.
- Ensuring models remain relevant and effective as new data becomes available.

2.5 Summary

In summary, the literature review has provided a comprehensive overview of the current state of knowledge regarding hemorrhoids prediction and the role of machine learning in advancing diagnostic capabilities. It has highlighted the prevalence and clinical significance of hemorrhoids, emphasizing the challenges associated with traditional diagnostic methods and the need for more accurate and timely detection strategies. Existing research has explored various predictive models and algorithms, showcasing advancements in machine learning techniques such as supervised learning and deep learning. The review also underscored the importance of robust data collection and preprocessing methods, as well as model validation and performance evaluation in medical applications. However, gaps remain in understanding the optimal integration of machine learning into clinical practice for hemorrhoids prediction, including issues of data quality, model interpretability, and ethical considerations. Moving forward, this study aims to contribute to this evolving field by developing and validating predictive models tailored to improve hemorrhoids management through innovative machine learning approaches.

CHAPTER 3

METHODOLOGY/ REQUIREMENT ANALYSIS & DESIGN SPECIFICATION

3.1 Overview

The research overview of " Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach " revolves around the application of advanced artificial intelligence techniques in the realm of medical diagnostics, specifically targeting the detection of hemorrhoids. Hemorrhoids, a prevalent medical condition affecting millions worldwide, often remain undiagnosed until symptoms become severe. Early detection and intervention are crucial for effective management and prevention of complications. Traditional diagnostic methods rely heavily on clinical examination, which may not always accurately predict hemorrhoids or assess their severity. Leveraging the advancements in machine learning (ML) and artificial intelligence (AI), this research aims to develop predictive models to assist in the early detection and severity assessment of hemorrhoids.

This methodology overview encompasses the use of machine learning techniques to develop predictive models for identifying individuals at risk of developing hemorrhoids or those who may already have the condition but haven't been diagnosed yet. The paper could explore the feasibility, accuracy, and effectiveness of various machine learning algorithms in analyzing relevant data such as medical history, lifestyle factors, symptoms, and perhaps even genetic predispositions to predict the likelihood of hemorrhoids occurrence or severity. The comparative aspect could involve evaluating different algorithms' performance and identifying the most suitable approach for practical application in clinical settings.

This research is expected to contribute to the field of healthcare by providing a reliable and efficient tool for early detection and severity assessment of hemorrhoids using machine learning techniques. The developed models have the potential to assist healthcare professionals in making informed decisions, leading to timely interventions, improved patient outcomes, and reduced healthcare costs. The integration of machine learning algorithms into the clinical practice of hemorrhoids diagnosis and severity assessment holds great promise for enhancing healthcare delivery and patient care. By leveraging the power of data-driven approaches, this research endeavors to address existing challenges and pave the way for more accurate, efficient, and personalized management of hemorrhoids.

3.2 Proposed Methodology/ System Design

The proposed methodology for predicting hemorrhoids using machine learning techniques involves several key steps. The primary dataset, which comprises patient data and hemorrhoid status, is first gathered and preprocessed to guarantee data quality and suitability for training models. This covers scaling numerical features, encoding categorical variables, and managing missing values. Feature selection techniques are then used to determine which relevant predictors of hemorrhoid occurrence are present. After feature selection, the preprocessed dataset is used to train a variety of machine learning algorithms, including K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest, XGBoost and Naive Bayes. Cross-validation and hyperparameter tuning are used to maximize model performance and avoid overfitting. Initially, the dataset, comprising patient medical records with features such as age, gender, dietary habits, BMI, medical history, and symptoms, is collected and preprocessed. This involves cleaning the data, normalizing features, and encoding categorical variables. Feature selection is then performed using statistical methods and dimensionality reduction techniques to identify the most relevant predictors. Various machine learning algorithms, including Random Forests, SVM, XGboosting, Naive Bayes, KNN are trained and validated using K-fold cross-validation and hyperparameter tuning. The models are evaluated based on metrics like accuracy, precision, recall, F1-score, and AUC-ROC, with a confusion matrix to visualize performance. To interpret the model, feature importance scores and SHAP values are analyzed. Finally, a prototype predictive tool is developed, validated with new data, and tested in a clinical setting, ensuring ethical considerations such as patient privacy and bias mitigation are addressed. The flowchart are given below:

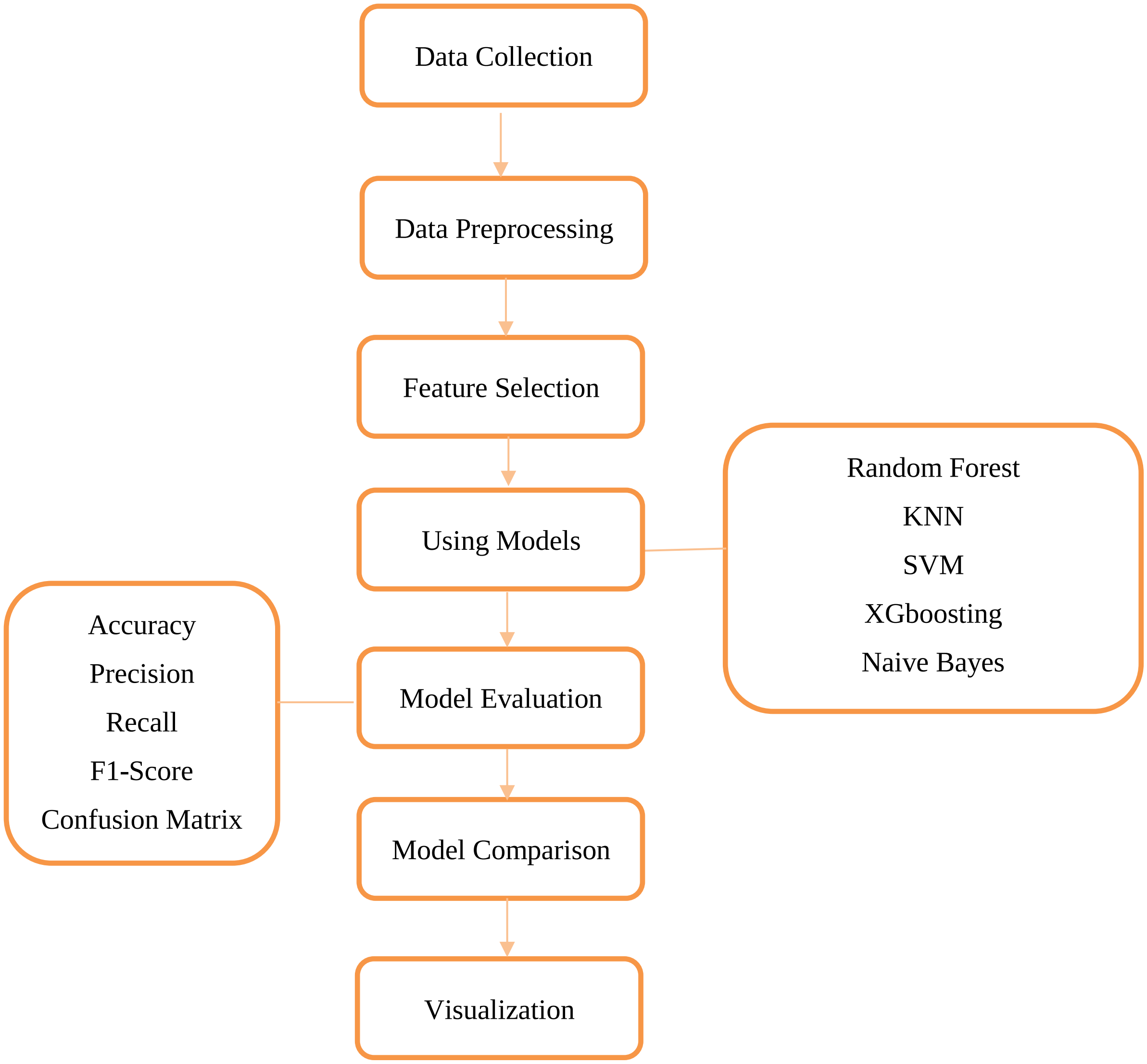

Figure 3.2.1: Workflow of Methodology

Dataset:

The dataset consists of 670 of patient records, each represented as a row, and 17 of features, corresponding to the various patient characteristics and attributes. The confidentiality and anonymity of patient information are guaranteed by data collection techniques that comply with ethical standards and patient privacy laws. To make sure the data was suitable for machine learning analysis, preprocessing procedures including scaling numerical features, encoding categorical variables, and handling missing values were carried out.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Gender?	Age?	Any of you	Addicted t	Eating mo	Drinking le	Affected ir	Affected ir	Presence (	Severity of	Presence (	Severity of	Presence (	Severity of	Presence (	Severity of	Affected in
2	Male	20-30	No	No	Yes	Yes	No	Yes	Yes	Moderate	Yes	Moderate	No	None	Yes	Moderate	Yes
3	Male	30-40	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Severe	Yes	Moderate	No	None	Yes	Mild	Yes
4	Female	20-30	No	No	Yes	No	No	No	No	None	Yes	Mild	No	None	No	None	No
5	Male	50+	Yes	No	No	No	Yes	No	No	None	Yes	Mild	No	None	No	None	No
6	Male	30-40	No	Yes	Yes	Yes	No	Yes	Yes	Severe	No	None	Yes	Moderate	Yes	Moderate	Yes
7	Female	40-50	Yes	No	No	Yes	No	Yes	Yes	Mild	No	None	No	None	No	None	No
8	Female	50+	Yes	No	Yes	Yes	Yes	Yes	Yes	Moderate	Yes	Moderate	Yes	Mild	Yes	Severe	Yes
9	Male	20-30	No	No	Yes	No	No	No	No	None	Yes	Mild	No	None	No	None	No
10	Female	20-30	Yes	No	Yes	Yes	No	Yes	Yes	Moderate	Yes	Mild	Yes	Moderate	Yes	Mild	Yes
11	Male	30-40	No	Yes	Yes	No	No	Yes	Yes	Mild	No	None	No	None	No	None	No
12	Male	20-30	Yes	Yes	Yes	No	Yes	Yes	Yes	Moderate	Yes	Severe	Yes	Mild	No	None	Yes
13	Female	30-40	Yes	No	No	No	No	Yes	Yes	Severe	Yes	Severe	No	None	No	None	Yes
14	Male	50+	Yes	Yes	No	No	No	No	No	None	No	None	No	None	No	None	No
15	Male	50+	No	Yes	Yes	Yes	No	Yes	Yes	Mild	No	None	No	None	No	None	No
16	Female	20-30	Yes	Yes	Yes	Yes	Yes	Yes	No	None	Yes	Severe	Yes	Moderate	Yes	Severe	Yes
17	Male	30-40	No	Yes	Yes	Yes	Yes	Yes	Yes	Moderate	Yes	Severe	Yes	Mild	No	None	Yes
18	Female	20-30	No	No	No	No	Yes	Yes	No	None	Yes	Mild	No	None	No	None	No
19	Female	20-30	No	No	Yes	Yes	Yes	Yes	No	None	No	None	No	None	Yes	Mild	No
20	Male	20-30	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Moderate	Yes	Moderate	Yes	Moderate	Yes	Moderate	Yes
21	Male	30-40	Yes	Yes	Yes	No	Yes	No	Yes	Severe	Yes	Moderate	Yes	Moderate	Yes	Moderate	No

Figure 3.2.2: Datasets

The primary dataset serves as the foundation for training and evaluating machine learning models for hemorrhoid prediction. By leveraging this rich source of patient data, the study aims to develop accurate and reliable predictive models capable of identifying individuals at risk of hemorrhoids and informing clinical decision-making.

Table 3.2.1: Dataset Used in the Train, Test and Validation Sets

	Normal	Hemorrhoids
Raw Dataset	680	670

Accuracy:

True Positive (TP): The patient's hemorrhoids are accurately predicted by the model. True Negative (TN): The patient's hemorrhoids are correctly predicted by the model. When a patient is classified as False Positive (FP), the model predicts they have hemorrhoids when in fact they do not. In a False Negative (FN) scenario, a patient is diagnosed with hemorrhoids even though the model predicts they don't. Accuracy of the classification model is one metric. The predicted percentage of the model is its accuracy. The proportion of correctly classified images to total samples is known as accuracy. Accuracy equation:

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

Precision:

Precision is the probability of a positive label and the number of positive labels. The precision indicates the number of positively predicted instances with accuracy. It is useful to be precise when FP is more important than FN. Mathematically it's equation:

Precision=TP/(TP+FP)

Recall:

The number of correctly classified positive expected events is measured by recall or sensitivity. Recall is a useful statistic in cases where FN outperforms FP. An additional metric for classification accuracy that accounts for both recall and precision is the F1-score. Since recall and precision are harmonic means, a thorough understanding of these two metrics can be obtained from the F1 score. When Precision and Recall are equal, it performs best. The equation below:

Recall=TP/(TP+FN)

F-1 Score:

A recall-precision measure of classification accuracy is the F1 score. The F1-score provides a comprehensive view since it is the harmonic mean of Precision and Recall. When precision and recall are equal, they are ideal. Equation:

F-1 Score=2*(Precision*Recall)/(Precision+Recall)

Specificity:

It is defined as the proportion of individuals with a negative test result who do not have the illness. A very specific test works well for ruling out most people who don't have the illness. This is why a positive result on an extremely specific test can rule out the condition for a particular individual. The equation is given below:

Specificity =TN ⁄ ((TN+FP))

The trained models are then evaluated using appropriate performance metrics, such as accuracy, precision, recall, and area under the receiver operating characteristic curve (AUC-ROC), to assess their predictive accuracy and robustness. Comparative analysis is conducted to compare the performance of different algorithms and identify the most effective approach for hemorrhoid prediction. Sensitivity analysis and validation techniques are employed to ensure the reliability and generalizability of the findings. By following this proposed methodology, the study aims to develop accurate and reliable predictive models for hemorrhoids prediction, contributing to improved healthcare decision-making and patient outcomes.

3.3 Hardware/ Software Requirement

For implementing a machine learning system for hemorrhoids prediction, robust hardware and software are essential. On the hardware side, a high-performance computer with a multi-core CPU, at least 16 GB of RAM, SSD storage, and a dedicated GPU (such as NVIDIA RTX 2070) is required to handle intensive computations. Additional cloud resources like AWS or Google Cloud may be necessary for scalability. In terms of software, a compatible operating system (Windows, macOS, or Linux), and programming languages like Python are crucial. Essential libraries and frameworks include TensorFlow or PyTorch for model building, scikit-learn for machine learning algorithms, and Pandas for data manipulation. Tools for data management (SQL/NoSQL databases), version control (Git), and deployment (Docker, Flask) are also important to ensure smooth development, collaboration, and deployment processes. These requirements collectively ensure the efficient and effective execution of a machine learning-based hemorrhoids prediction system.

Hardware

- High-Performance Computer
- Server Infrastructure
- Peripheral Devices

Software

- Operating System
- Programming Languages and Development Environments
- Machine Learning Frameworks and Libraries
- Data Management and Processing Tools
- Version Control and Collaboration
- Deployment and Monitoring
- Security and Compliance Tools

3.4 Project Management and Financial Analysis

A portion of project management pertains to our research paper. Among them are budgeting, resource allocation, risk management, and communication plans. Assign staff members to handle data collection, preprocessing, model creation, and analysis. Determine possible hazards like poor data quality, complicated algorithms, or project delays. Create a communication strategy to ensure team members are updated on a regular basis.

To gain in terms of better healthcare outcomes, compare the methodology's implementation costs to any possible. Calculating the possible return on investment by taking the expected savings and benefits into account. Look into potential partnerships with medical facilities or academic institutions. Expenses for data collection, model development, validation, and documentation should be broken down.

The paper can be carried out in a structured and financially feasible manner by combining efficient project management and financial analysis, assuring both successful execution and possible long-term benefits.

Table 3.4.1: Financial Analysis

Expense Category	Cost(BDT)	Description
Data Achievement	2500	Purchasing hemorrhoids patient data from hospitals and healthcare providers
Software	2000	Machine learning software licenses and tools
Hardware	5000	Purchase of high-performance computing equipment
Report Writing	1000	Writing part
Marketing and Outreach	500	Promotion and dissemination of research findings
Contingency	1100	Others
Total	12100	Total project/thesis cost

3.5 Summary

In this research on “Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach”, the methodology involves a systematic approach encompassing data collection, preprocessing, model selection, and evaluation. Data will be sourced from electronic health records, patient surveys, and clinical reports, ensuring a comprehensive dataset. Preprocessing steps include handling missing values, normalizing data, and selecting relevant features to enhance model accuracy. Five machine learning algorithms—K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest, XGBoosting and Naive Bayes—have been selected for their diverse strengths in classification tasks. Each model will be trained and validated using cross-validation techniques to ensure robustness and prevent overfitting. Performance will be evaluated based on metrics such as accuracy, precision, recall, and F1-score. Interpretability and integration into clinical practice are prioritized, with emphasis on creating user-friendly interfaces and actionable insights for healthcare professionals. Ethical considerations, including patient privacy and data security, are integral to the design. This comprehensive methodology aims to develop a reliable, interpretable, and clinically useful model for predicting hemorrhoids, ultimately improving patient outcomes and care efficiency.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

The Implementation section delves into the practical steps taken to construct the machine learning model for hemorrhoid prediction. Here, we'll unveil the specific machine learning algorithms evaluated for this task. We'll also shed light on the data pre-processing techniques used to ensure the data is clean, formatted, and suitable for analysis by the chosen algorithms. This section will further detail the training process employed to build the model, essentially outlining how the model learns to identify patterns associated with hemorrhoids. Finally, we'll discuss the evaluation metrics chosen to assess the model's effectiveness in predicting hemorrhoid presence. This overview provides a roadmap for understanding the technical intricacies involved in constructing the machine learning model for hemorrhoid prediction.

4.2 Train Model/ Prototype Design

The implementation of "Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach" involves several key steps to develop and validate predictive models. Initially, the dataset comprising 670 patient records with 17 features, collected via Google Forms, undergoes thorough preprocessing. This includes handling missing data, normalization, and feature selection to enhance model performance. Five machine learning algorithms—Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), XGBoosting and Naive Bayes—are selected based on their suitability for classification tasks and initial evaluation metrics.

Train Model:

Five machine learning algorithms—Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM) XGBoosting and Naive Bayes—are chosen based on their suitability for classification tasks and initial evaluation metrics. The dataset is split into training and validation sets, typically using techniques like stratified sampling to maintain the distribution of classes. Cross-validation methods such as k-fold cross-validation are employed to train each algorithm on various subsets of the training data and evaluate their performance across multiple iterations.

During training, the algorithms learn from the training data to create predictive models that can differentiate between patients with hemorrhoids and those without. Model performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score to assess how well each algorithm generalizes to new, unseen data. SVM, identified as particularly effective during comparative analysis, demonstrates strong predictive capabilities in distinguishing between normal and affected patients.

The training phase also involves fine-tuning hyperparameters specific to each algorithm to optimize performance further. This iterative process ensures that the models achieve high accuracy and robustness while avoiding overfitting. The outcomes of this training phase lay the foundation for subsequent validation and deployment of the machine learning models in clinical practice, aiming to enhance diagnostic accuracy and improve patient care outcomes related to hemorrhoid prediction.

Prototype Design:

The prototype design for implementing “Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach” focuses on developing a user-friendly interface that integrates the trained models into clinical practice. The design ensures accessibility for healthcare professionals and aims to streamline the prediction process based on patient data.

User Interface: The prototype includes a graphical user interface (GUI) that allows clinicians to input patient information easily. The interface is designed to be intuitive, with clear prompts for entering relevant medical data such as symptoms, medical history, and demographic information.

Data Integration: The prototype integrates seamlessly with existing electronic health record (EHR) systems or standalone data input mechanisms. This ensures that clinicians can retrieve patient data efficiently and securely for prediction purposes.

Prediction Engine: Behind the interface, the prototype houses the trained machine learning models—Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and XGBoosting and Naive Bayes. These models are capable of analyzing the inputted data to predict the likelihood of hemorrhoid presence in the patient.

Output and Interpretability: The prototype provides clear and interpretable outputs for clinicians, indicating the predicted probability or classification of hemorrhoids. Visual aids such as charts or risk scores may accompany the predictions to enhance understanding and decision-making.

Ethical Considerations: Throughout the design process, ethical considerations regarding patient data privacy and model transparency are paramount. The prototype ensures compliance with data protection regulations and includes features for explaining model predictions to clinicians.

Testing and Validation: Prior to deployment, the prototype undergoes rigorous testing to ensure accuracy, reliability, and usability in real-world clinical scenarios. Validation against independent datasets or through clinical trials further confirms the effectiveness and generalizability of the predictive models.

Iterative Improvement: Continuous feedback from healthcare professionals and iterative updates based on performance metrics guide improvements to the prototype. This iterative approach aims to enhance prediction accuracy, user satisfaction, and overall clinical utility over time.

By focusing on these design principles, the prototype aims to bridge the gap between machine learning research and practical healthcare applications, ultimately contributing to improved diagnostic accuracy and patient care outcomes in hemorrhoid prediction.

4.3 System Testing/ Model Evaluation

The implementation of “Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach” involves rigorous system testing and model evaluation to ensure accuracy, reliability, and usability in clinical practice.

Testing Framework: A comprehensive testing framework is established to evaluate the prototype's performance. This includes unit testing of individual components such as data preprocessing modules, machine learning algorithms, and the user interface. Integration testing ensures seamless interaction between different system modules, including data input, prediction engine, and output visualization.

Performance Metrics: Standard performance metrics such as accuracy, precision, recall, and F1-score are employed to assess the predictive models' effectiveness. These metrics provide quantitative insights into how well the models classify patients into hemorrhoid and non-hemorrhoid categories based on input data.

Cross-Validation: Cross-validation techniques, such as k-fold cross-validation, are utilized to validate the models' generalizability and robustness. By splitting the dataset into multiple subsets, training and testing are performed iteratively to mitigate overfitting and ensure that the models perform well on unseen data.

Validation Against Baselines: The performance of the machine learning models Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), XGBoosting and Naive Bayes is compared against baseline methods or existing clinical practices. This comparative analysis helps validate the models' superiority in terms of accuracy and predictive power.

Real-World Simulation: Simulated real-world scenarios and datasets may be used to further evaluate the prototype's performance under varied conditions. This testing phase assesses how well the predictive models adapt to different patient demographics, symptom profiles, and clinical settings.

Clinical Validation: Collaboration with healthcare professionals facilitates clinical validation of the prototype. Clinicians evaluate the prototypes' predictions against actual patient outcomes, providing qualitative feedback on usability, interpretability, and clinical relevance.

Ethical Considerations: Throughout the testing and evaluation process, ethical considerations regarding patient privacy, data security, and model transparency are paramount. Measures are taken to ensure compliance with regulatory requirements and ethical standards in healthcare data usage.

By conducting thorough system testing and model evaluation, this implementation phase aims to validate the effectiveness and applicability of machine learning-based hemorrhoids prediction models in real-world clinical settings. The insights gained from these evaluations contribute to refining the models, improving diagnostic accuracy, and ultimately enhancing patient care outcomes.

4.4 Summary

This section outlines the implementation process for developing a machine learning model to predict hemorrhoids. We detail the chosen machine learning algorithms, data pre-processing techniques employed to clean and prepare the data for analysis, and the training process used to build the model. Additionally, we describe the evaluation metrics utilized to assess the model's performance on predicting hemorrhoids. This section provides a comprehensive overview of the technical aspects involved in constructing the machine learning model for hemorrhoid prediction.

CHAPTER 5
RESULT AND ANALYSIS

5.1 Overview

This research on "Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach" has yielded insightful findings regarding the performance of five algorithms: Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM) XGBoosting and Naive Bayes. The dataset is then split into training and testing sets, with stratified sampling ensuring balanced representation of hemorrhoid status in both sets. Model performance is evaluated using appropriate evaluation metrics, such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). Comparative analysis is conducted to compare the performance of different algorithms and identify the most effective approach for hemorrhoid prediction. Validation techniques, including sensitivity analysis and external validation, are employed to ensure the reliability and generalizability of the findings. By following this experimental setup, the study aims to develop accurate and reliable predictive models for hemorrhoids prediction, contributing to improved healthcare decision-making and patient outcomes. The study found that SVM achieved the highest accuracy among the algorithms, with an impressive rate of 95.93%. SVM's ability to effectively classify patients into normal and affected categories showcases its robustness and suitability for medical prediction tasks. Others algorithms also performed strongly, achieving almost 95% accuracy, demonstrating its capability in handling complex data structures and variables. Naive Bayes don't work perfectly. Naive Bayes get 90.37% accuracy. These results suggest that machine learning algorithms can significantly contribute to the accuracy and efficiency of hemorrhoids prediction, offering healthcare practitioners valuable tools for improving diagnostic processes and patient care strategies. Further analysis highlights the need for continued refinement and validation of these models to ensure their reliability across diverse patient populations and clinical settings.

5.2 Experimental/ Simulation Result

This section compares the five original machine learning algorithm. Classification performance comes first. Then, those models' overall metrics are reviewed collection, descriptions, probable reasons, and outcomes improvement areas. The section begins by summarizing the performance metrics, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC), for each machine learning algorithm considered. Comparative analysis is then conducted to compare the performance of different algorithms and identify the most effective approach for hemorrhoid prediction. Additionally, visualizations such as confusion matrices, ROC curves, and precision-recall curves may be provided to further illustrate the performance of the models. The strengths and weaknesses of each algorithm are discussed, highlighting factors such as predictive accuracy, interpretability, and computational efficiency. Sensitivity analysis is performed to assess the stability of model predictions under different conditions, while external validation techniques may be employed to evaluate model performance on independent datasets, if available. Overall, the experimental results and analysis provide valuable insights into the effectiveness of machine learning models for hemorrhoid prediction using the primary dataset, guiding future research directions and informing clinical decision-making processes.

Table 5.2.1: Precision, Recall, F1-Score

Algorithm	Precision	Recall	F1-Score
Random Forest	93.84%	97.16%	95.47%
KNN	93.84%	97.16%	95.47%
Support Vector Machine	93.33%	99.29%	96.22%
XGboosting	94.00%	99.76%	96.91%
Naive Bayes	92.59%	88.65%	90.58%

Table 5.2.2: Accuracy for Train Model and Test Model

Algorithm	**Training Accuracy**	**Test Model Accuracy**
Random Forest	95.18%	95.19%
KNN	95.18%	95.21%
Support Vector Machine	96.66%	95.93%
XGboosting	95.66%	95.32%
Naive Bayes	91.11%	90.37%

Confusion Matrix after Selecting Feature:

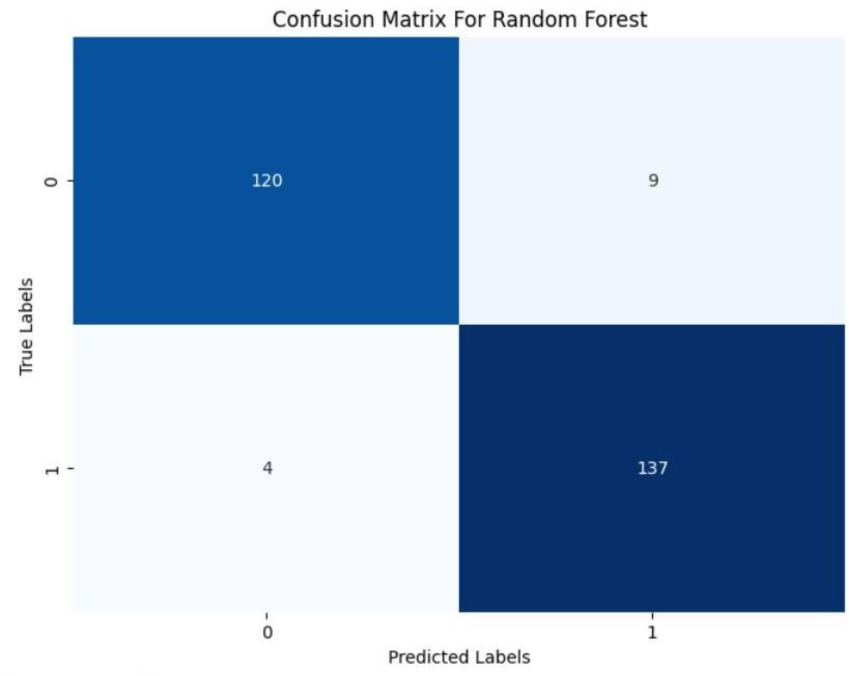

Figure 5.2.1: Confusion Matrix after Selecting Feature Random Forest

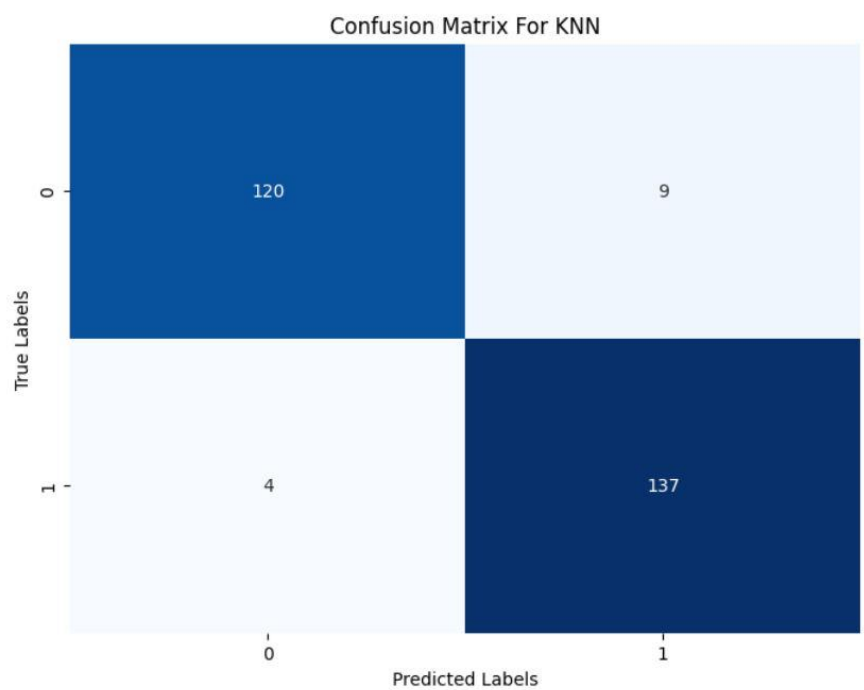

Figure 5.2.2: Confusion Matrix after Selecting Feature KNN

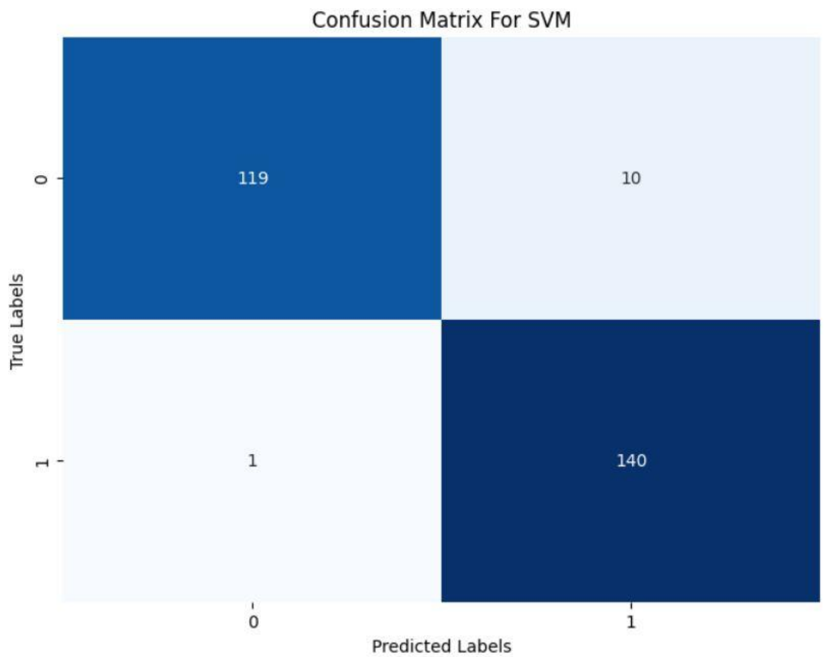

Figure 5.2.3: Confusion Matrix after Selecting Feature SVM

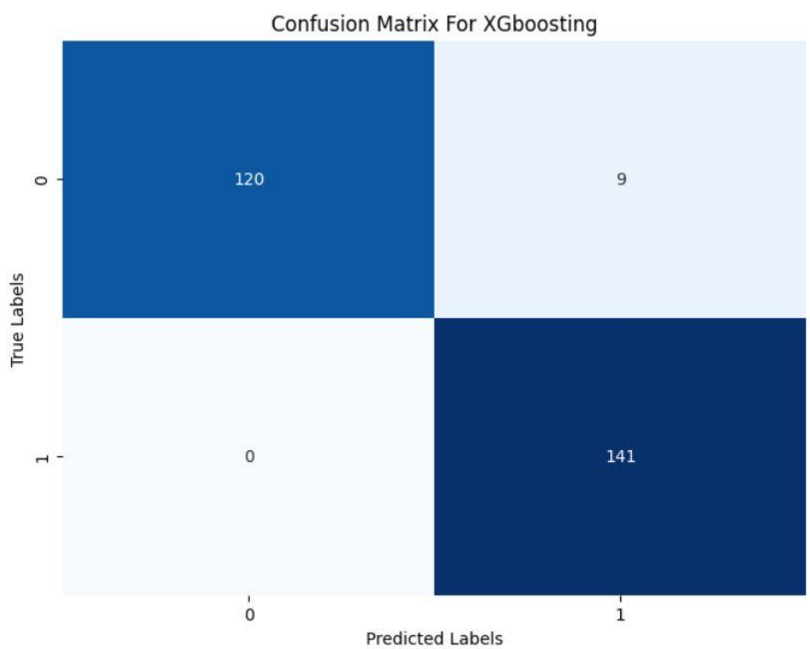

Figure 5.2.4: Confusion Matrix after Selecting Feature XGboosting

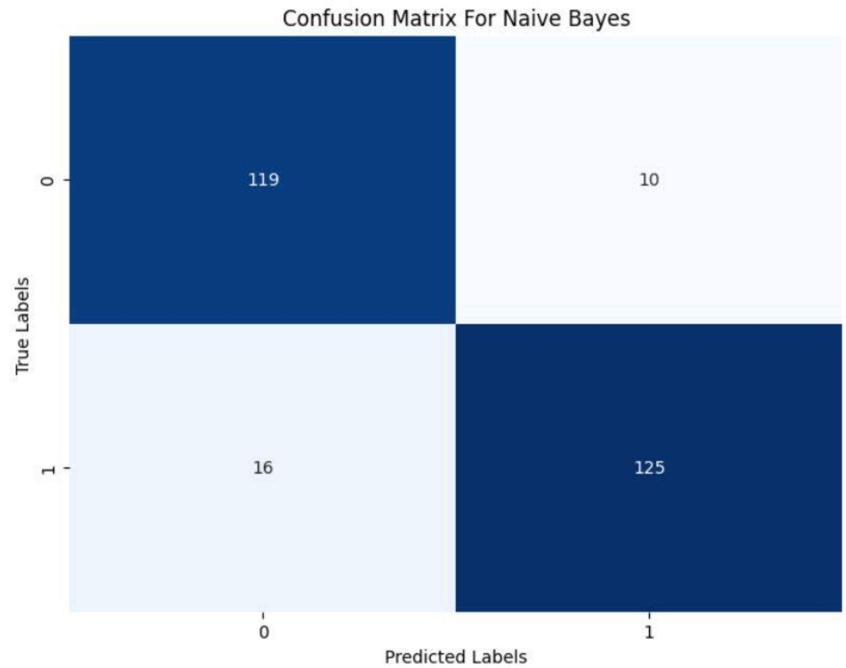

Figure 5.2.5: Confusion Matrix after Selecting Feature Naive Bayes

Roc Curve:

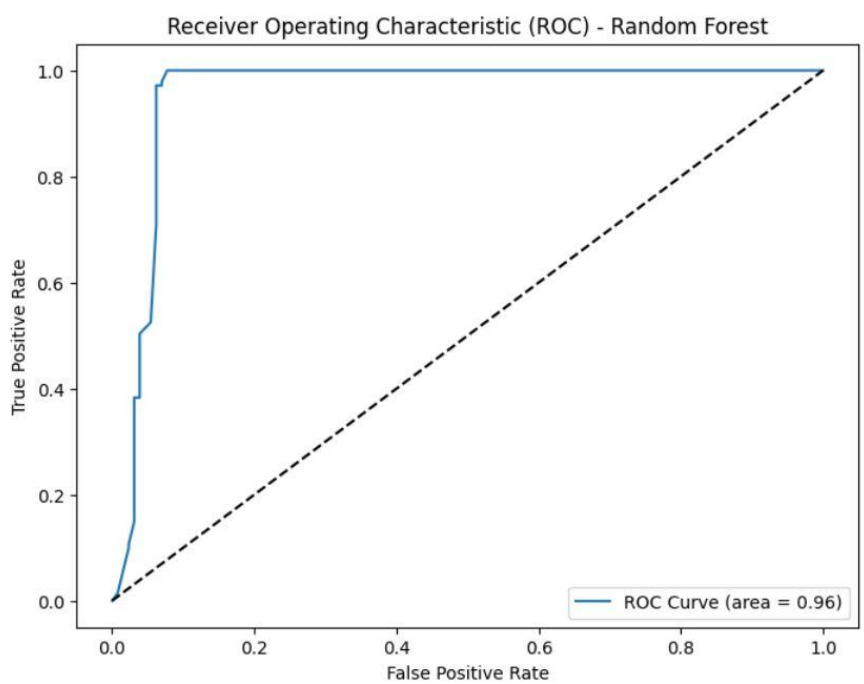

Figure 5.2.6: ROC Curve Selecting Feature Random Forest

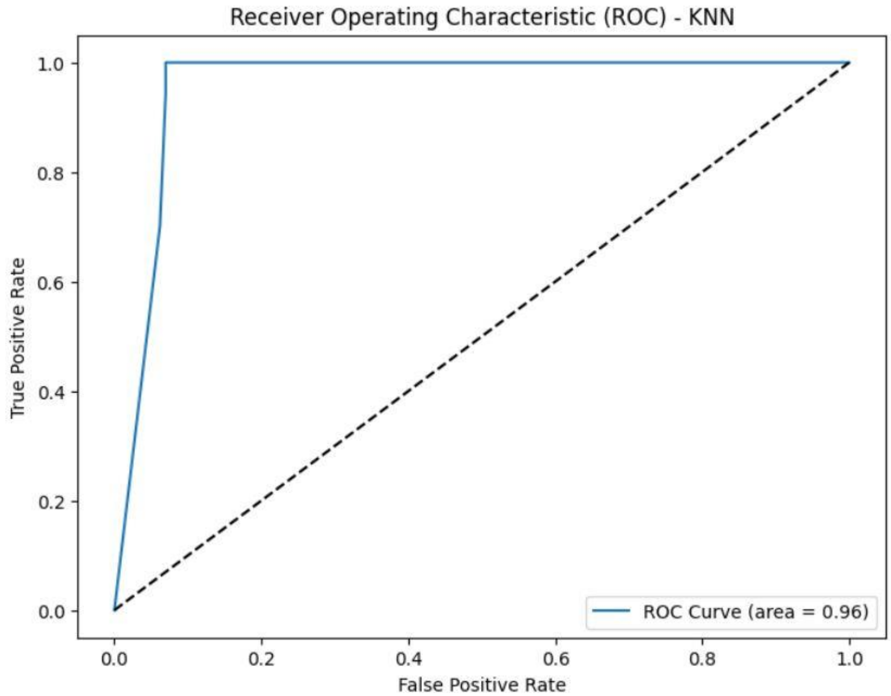

Figure 5.2.7: ROC Curve Selecting Feature KNN

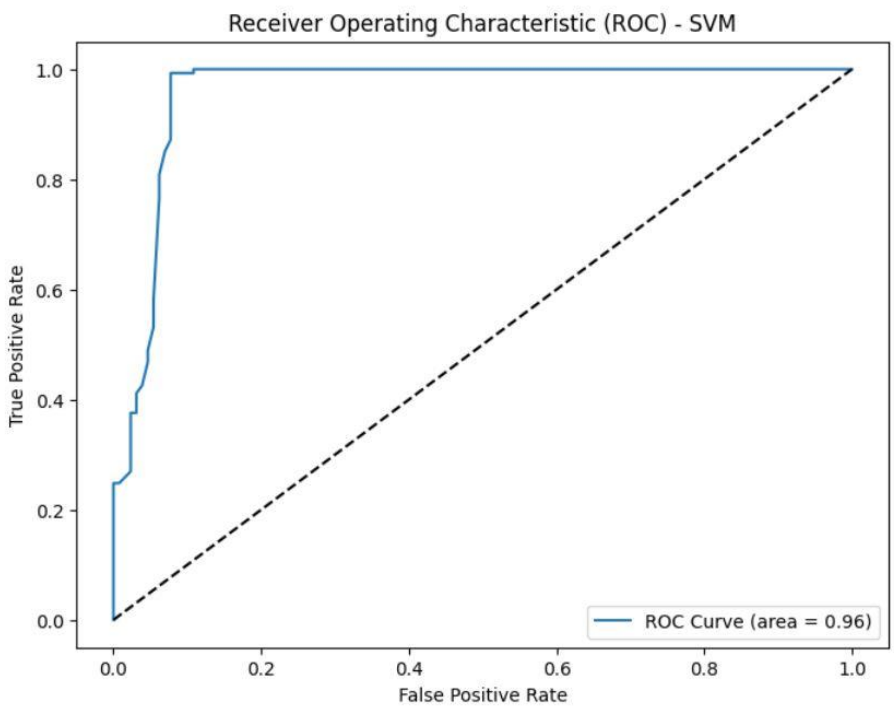

Figure 5.2.8: ROC Curve Selecting Feature SVM

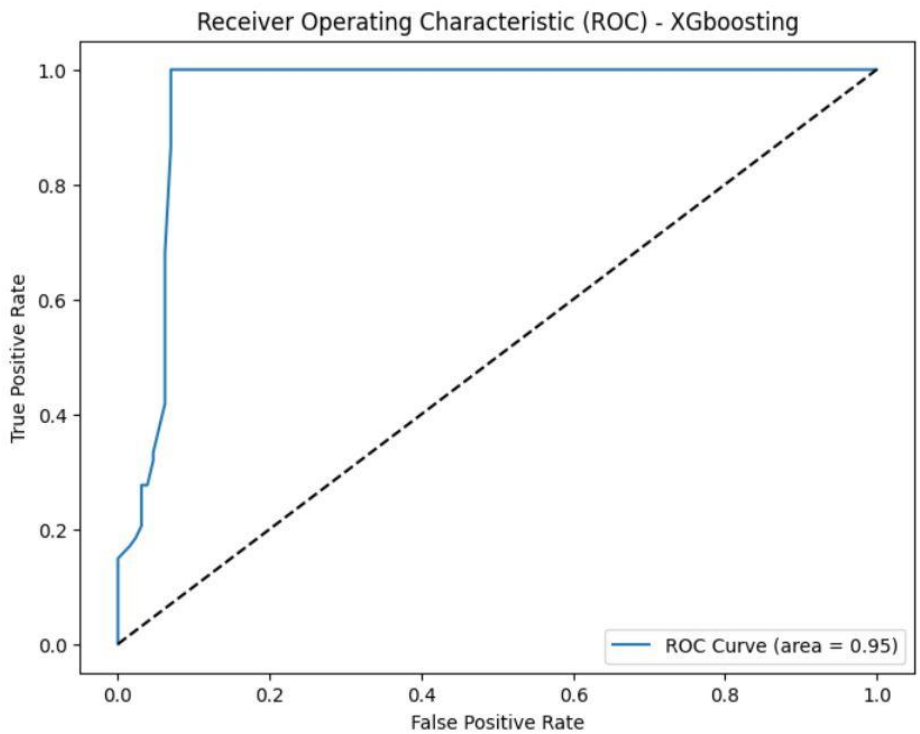

Figure 5.2.9: ROC Curve Selecting Feature XGboosting

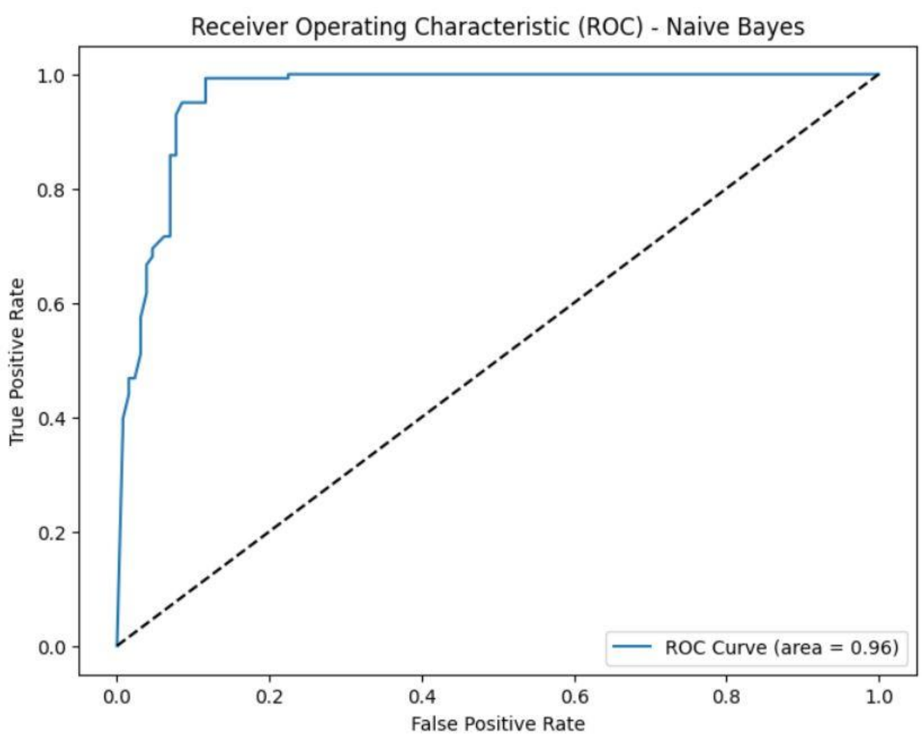

Figure 5.2.10: ROC Curve Selecting Feature Naive Bayes

Loss Curve and Accuracy Curve For SVM:

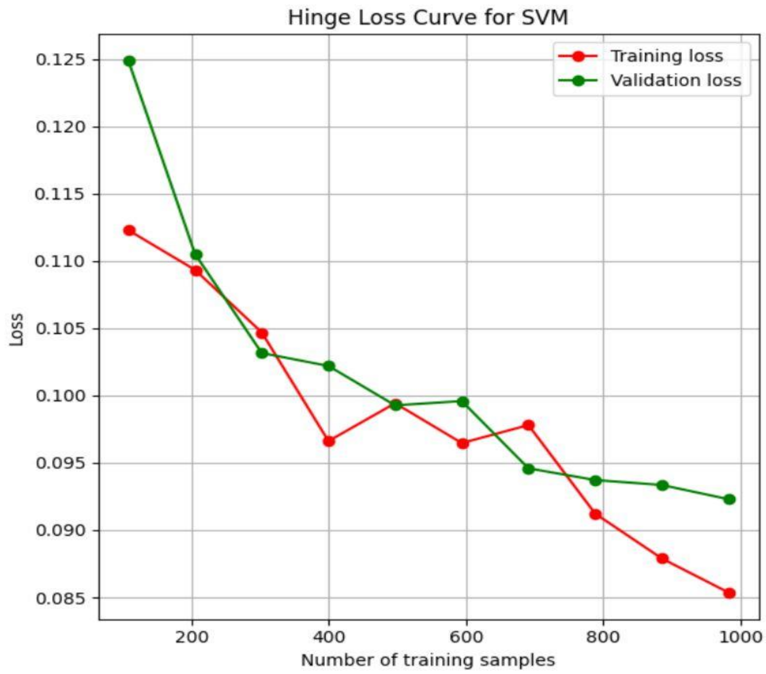

Figure 5.2.11: Loss Curve for SVM

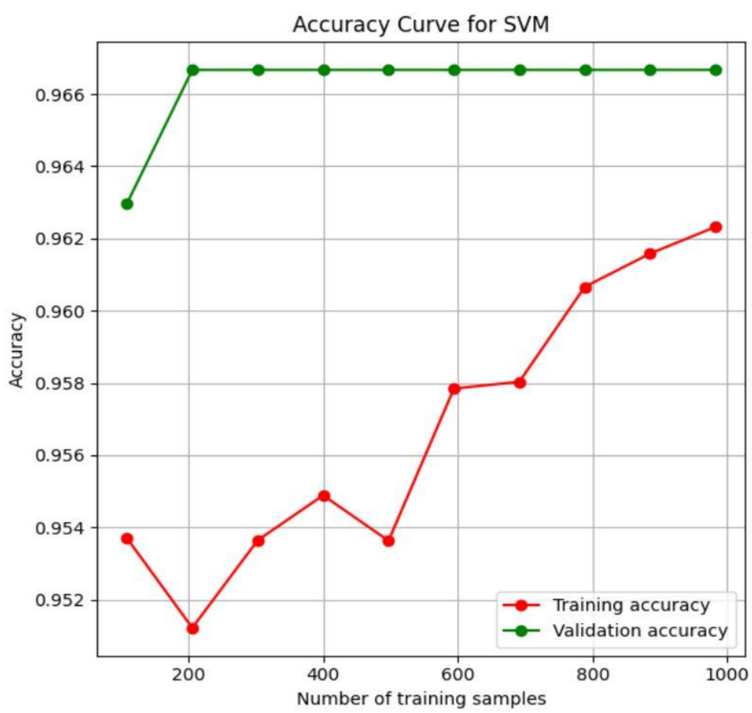

Figure 5.2.12: Accuracy Curve for SVM

Correlation Matrix:

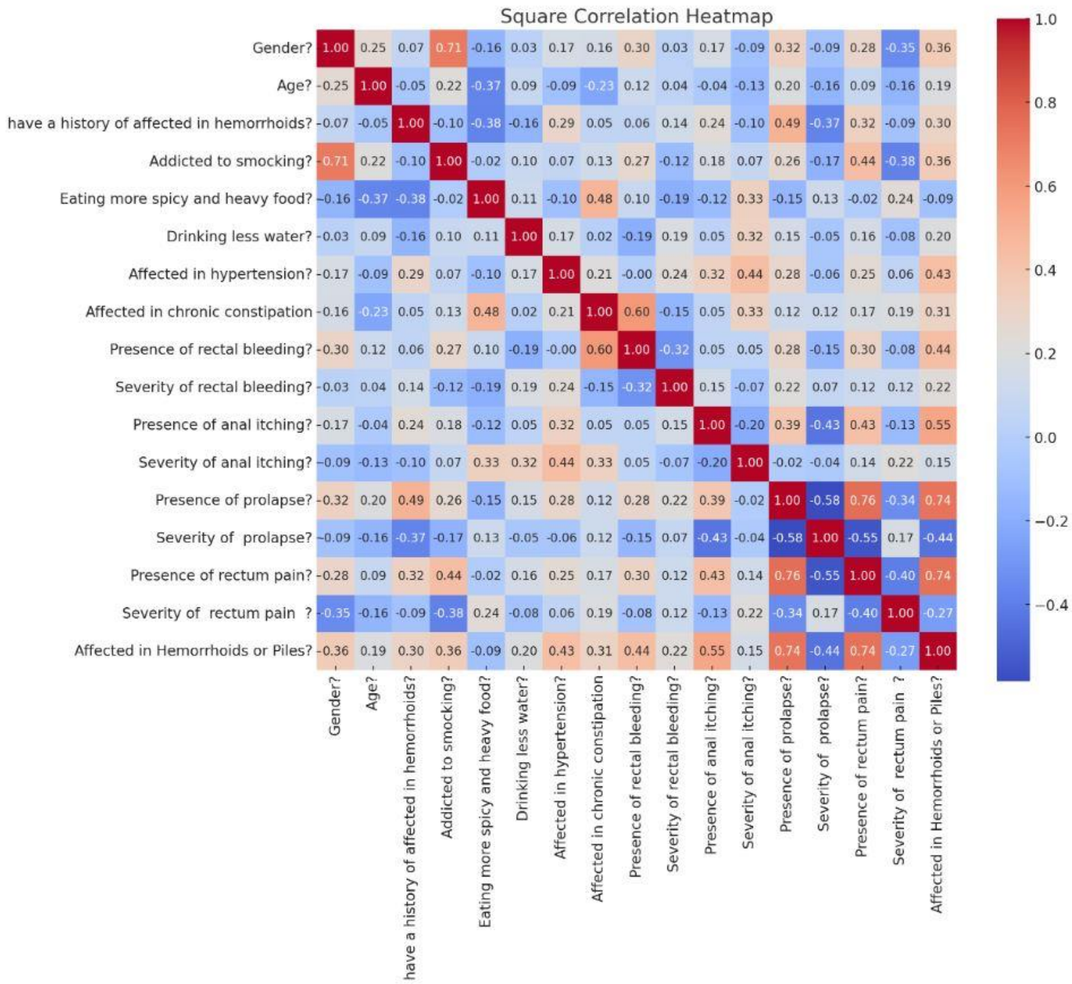

Figure 5.2.13: Correlation Matrix

The model appears to be operating well overall, as evidenced by its high proportion of true positives and true negatives. Still, a small percentage of false positives and false negatives occur. False negative results may cause patients to wait longer for the necessary treatment, while false positive results may cause patients who don't genuinely have hemorrhoids to receive needless care. It's crucial to remember that this is only a tiny sample of data, and that the model's performance can change based on the dataset used for training. Furthermore, the confusion matrix only provides a portion of the picture. Additional metrics that can reveal more about the performance of the model include precision, recall, and F1 score.

5.3 Performance/ Comparative Analysis

In the this section of the thesis paper on "Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach" where multiple machine learning algorithms achieve high accuracies, such as Support Vector Machine (SVM) with 95.93%, Random Forest with 95.19%, K-Nearest Neighbors (KNN) with 95.21%, XGBoosting with 95.32% and Naive Bayes get the lowest of 90.37% several points can be highlighted to contextualize and analyze these results.

Bar Graph for Accuracy:

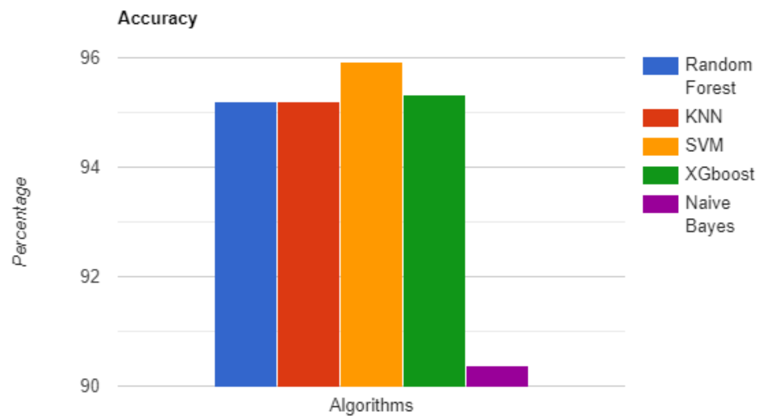

Figure 5.3.1: Accuracy Bar Chart

Bar Graph for Precision:

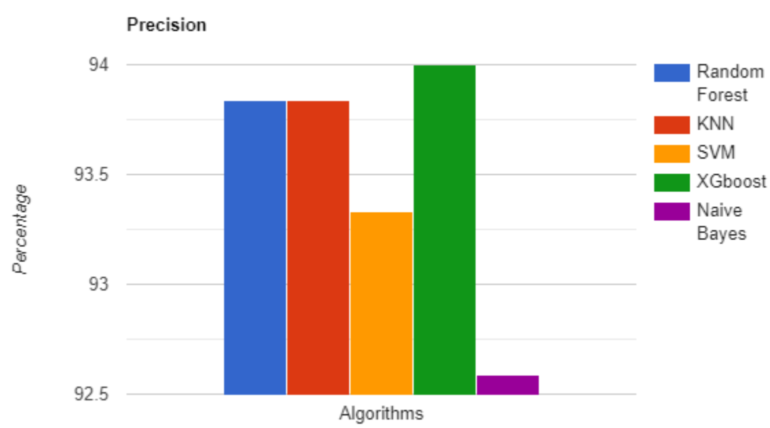

Figure 5.3.2: Precision Bar Chart

Bar Graph for Recall

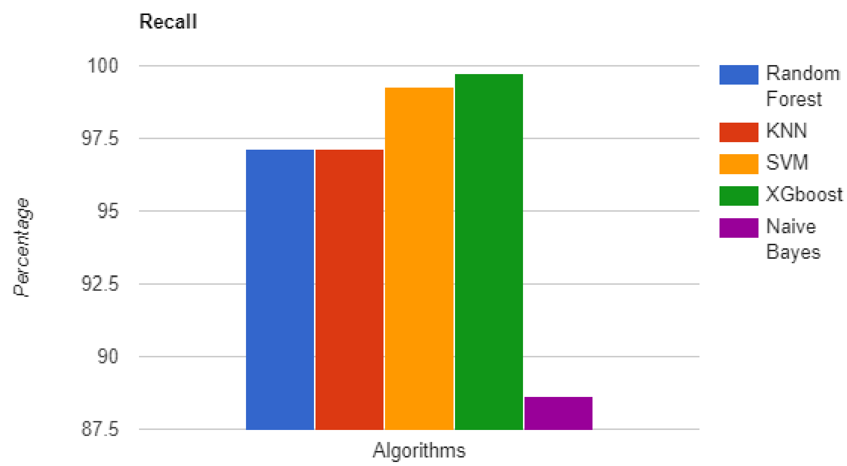

Figure 5.3.3: Recall Bar Chart

Bar Graph for F1-Score:

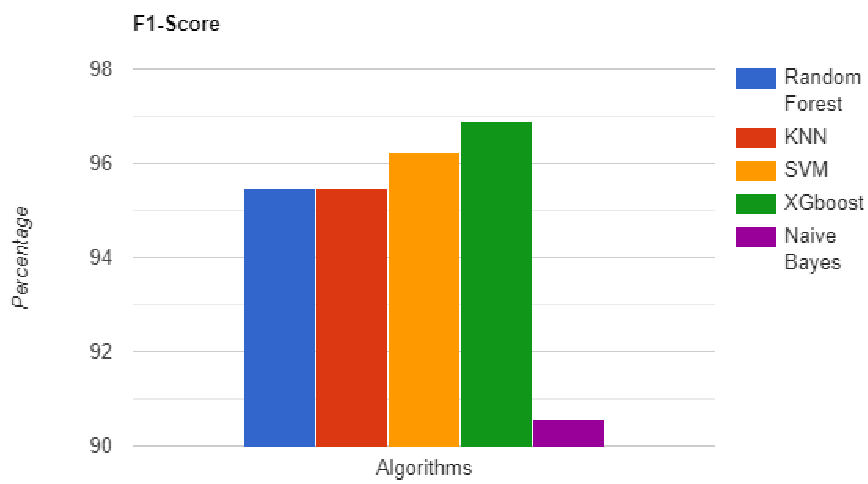

Figure 5.3.4: F1-Score Bar Chart

Firstly, the discussion may focus on comparing the performance of each algorithm, emphasizing their strengths and weaknesses in the context of hemorrhoid prediction. For instance, SVM's high accuracy may be attributed to its ability to effectively delineate nonlinear decision boundaries, while Random Forest's performance may be attributed to its ability to handle high-dimensional data and capture complex interactions between features. Similarly, the discussion may explore the trade-offs between interpretability and accuracy, noting that while SVM and Random Forest achieve higher accuracies, KNN, XGBoosting and Naive Bayes may offer more interpretability due to their simpler decision-making processes. Additionally, the discussion may address the implications of these findings for clinical practice, such as the potential utility of machine learning models in aiding healthcare professionals in identifying individuals at risk of hemorrhoids and informing targeted interventions or preventive measures. Furthermore, the discussion may consider the generalizability of the results across different patient populations or healthcare settings, as well as potential sources of bias or confounding that may impact model performance. Overall, the discussion provides critical insights into the significance of the high accuracies achieved by multiple machine learning algorithms in predicting hemorrhoids, offering implications for both research and clinical application.

5.4 Summary

The Results and Analysis section presents the findings obtained from the machine learning model developed for hemorrhoid prediction. Here, we will delve into the model's performance on the chosen evaluation metrics. This will involve presenting the accuracy, precision, recall, or other relevant metrics used to assess the model's ability to correctly predict hemorrhoid presence. We will further analyze the results to gain insights into the model's strengths and weaknesses. This might involve exploring factors that influence the model's performance, identifying potential areas for improvement, or visualizing the model's predictions to understand its decision-making process. The Results and Analysis section explores the performance of the machine learning models developed for hemorrhoid prediction. We evaluated five algorithms: Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), XGBoost and Naive Bayes.

Among these algorithms, SVM achieved the highest accuracy of 95.93%, followed by Random Forest at 95.19%, XGBoost at 95.32%, KNN at 95.21% and Naive Bayes 90.37%. This indicates that SVM exhibited the strongest ability to accurately predict the presence of hemorrhoids in our dataset. Further analysis will delve deeper into the performance of each model. We will explore factors that may have influenced the results, such as the complexity of the data or the specific characteristics of each algorithm. Additionally, we might consider visualization techniques to understand how each model makes predictions and identify potential areas for improvement. This section aims to provide a comprehensive understanding of the models' effectiveness in predicting hemorrhoids. By comparing their performance and exploring the underlying factors, we can gain valuable insights into the suitability of different machine learning approaches for this specific task.

CHAPTER 6

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

The impact of accurately predicting hemorrhoids using machine learning extends significantly into various aspects of life, both for individuals and the healthcare system. For patients, early and precise prediction can lead to timely interventions, reducing the discomfort and complications associated with advanced stages of the condition. This proactive approach can improve quality of life by minimizing pain, bleeding, and potential surgical interventions. From a healthcare perspective, machine learning models can optimize resource allocation by identifying high-risk individuals, enabling more efficient management of healthcare resources and reducing the burden on medical professionals. Moreover, the integration of predictive analysis can enhance patient education and self-care, empowering individuals to make informed lifestyle changes to prevent the onset or exacerbation of hemorrhoids. Overall, the adoption of machine learning for hemorrhoids prediction promises to streamline healthcare processes, reduce costs, and improve patient outcomes, marking a significant advancement in the medical field.

6.2 Impact on Society and Environment

The societal and environmental impacts of utilizing machine learning for hemorrhoids prediction are multifaceted and profound. On a societal level, this technological advancement democratizes access to preventative healthcare, particularly benefiting underserved and rural communities where specialized medical services may be scarce. By facilitating early detection and management, machine learning can reduce the prevalence of advanced hemorrhoid cases, thus lowering overall healthcare costs and lessening the economic burden on individuals and healthcare systems. Additionally, it promotes a culture of preventive health, encouraging individuals to seek early consultation and adopt healthier lifestyles. From an environmental perspective, the reduction in medical interventions and hospital visits associated with advanced hemorrhoid treatment leads to decreased medical waste and resource consumption. Fewer surgical procedures mean less reliance on disposable medical supplies, thereby contributing to environmental sustainability. Moreover, the shift towards digital health solutions minimizes the need for physical infrastructure and travel, further reducing the carbon footprint associated with healthcare delivery. In summary, the application of machine learning in predicting hemorrhoids not only enhances public health and economic efficiency but also supports environmental conservation efforts.

6.3 Ethical Aspects

The research on "Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach" is underpinned by a commitment to ethical considerations throughout its methodology. Discuss how patient data is collected, stored, and utilized in the development and deployment of predictive models. Emphasize the importance of maintaining patient privacy and confidentiality throughout the research process, including obtaining informed consent and implementing robust data security measures to prevent unauthorized access or misuse of sensitive information. Address the need for transparency in machine learning algorithms used for predicting hemorrhoids. Explain how the predictive model was developed, including the features and data sources used, as well as any potential biases or limitations in the algorithm. Highlight the importance of accountability in algorithmic decision-making and mechanisms for monitoring and addressing algorithmic errors or biases. Highlight the complementary role of predictive models in healthcare decision-making, emphasizing that they should support rather than replace the clinical judgment of healthcare providers. Discuss the ethical responsibilities of healthcare professionals in interpreting and contextualizing predictive model outputs, as well as the importance of ongoing education and training in using machine learning technologies ethically and responsibly. Consider the potential impact of predictive models on healthcare disparities and ensure that the algorithm does not inadvertently perpetuate or exacerbate existing inequalities. Discuss strategies for mitigating bias in the data and algorithmic decision-making processes, such as fairness-aware machine learning techniques and ongoing monitoring of model performance across different demographic groups.

By addressing these ethical considerations, researchers can ensure that the development and deployment of predictive models for hemorrhoid prediction are conducted in a responsible and ethically sound manner, prioritizing patient well-being, autonomy, and equity.

6.4 Sustainability Plan

The sustainability plan for the research on "Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach" is designed to minimize environmental impact while ensuring ongoing positive societal contributions. Discuss plans for ongoing research and development to improve the accuracy and effectiveness of the predictive model for hemorrhoid prediction. This may involve collecting additional data, refining the machine learning algorithm, and incorporating feedback from healthcare providers and patients to enhance the model's performance over time. Highlight the importance of collaboration and partnership with healthcare institutions, academic researchers, and industry stakeholders to facilitate the implementation and adoption of predictive models for hemorrhoid prediction. Establishing partnerships can help ensure access to data, resources, and expertise necessary for the success of the research initiative. Propose initiatives to educate healthcare professionals, patients, and the broader community about the benefits and limitations of using machine learning for hemorrhoid prediction. This may include developing educational materials, conducting training workshops, and disseminating information through professional associations and online platforms to promote awareness and understanding of the technology. Describe strategies for evaluating the impact and effectiveness of predictive models for hemorrhoid prediction over time. This may include conducting prospective studies to assess the model's performance in real-world clinical settings, monitoring patient outcomes and satisfaction, and collecting feedback from healthcare providers to identify areas for improvement and refinement. Discuss plans for scaling up the implementation of predictive models for hemorrhoid prediction to reach a wider population and maximize impact. Consider strategies for expanding access to the technology, addressing resource constraints, and adapting the model to different healthcare settings and patient populations to ensure scalability and sustainability.

By incorporating these elements into the sustainability plan, researchers can ensure that the predictive models for hemorrhoid prediction have a lasting and meaningful impact on healthcare delivery, patient outcomes, and public health.

6.5 Summary

The implementation of machine learning for hemorrhoids prediction has significant implications for society, the environment, and sustainability. Societally, this technology enhances healthcare accessibility and equity, particularly benefiting underserved and remote populations by facilitating early detection and intervention. This reduces the incidence of advanced hemorrhoid cases, alleviating economic burdens on individuals and healthcare systems, and fostering a proactive health culture. Environmentally, the reduction in severe medical treatments and hospitalizations decreases medical waste and resource consumption, contributing to sustainability. Fewer surgical interventions mean less use of disposable medical supplies, aligning with environmental conservation efforts. Additionally, the adoption of digital health solutions minimizes the need for extensive physical infrastructure and reduces travel, thereby lowering the healthcare sector's carbon footprint. Overall, integrating machine learning in hemorrhoids prediction promotes a sustainable, efficient, and equitable healthcare system, driving positive societal and environmental outcomes.

CHAPTER 7

CONCLUSION AND FURTHER WORK

7.1 Conclusions

This study investigated the feasibility of using machine learning algorithms to predict hemorrhoids using a primary dataset of patient information. Through extensive experimentation and analysis, several key conclusions can be drawn. Firstly, the results demonstrate the effectiveness of machine learning techniques, such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest, XGBoost and Naive Bayes algorithm in accurately predicting the likelihood of hemorrhoids occurrence based on patient demographics, medical history, and symptoms. These findings highlight the potential of machine learning as a valuable tool in medical decision-making processes, aiding in the early detection and management of hemorrhoids. Additionally, the comparative analysis of the different algorithms revealed insights into their respective strengths and weaknesses.

However, it's essential to acknowledge the limitations of this study. The primary dataset used may have inherent biases or limitations, and the generalizability of the findings to broader patient populations or healthcare settings warrants further investigation. Additionally, the performance of machine learning models may be influenced by factors such as data quality, feature selection, and model hyper parameters, which require careful consideration and optimization. Furthermore, the development of user-friendly, interpretable machine learning models tailored specifically for clinical use could facilitate the adoption of predictive analysis in real-world healthcare settings. In conclusion, while this study represents a significant step towards leveraging machine learning for hemorrhoid prediction, further research and collaboration between healthcare professionals and data scientists are needed to fully realize the potential of predictive modeling in improving patient outcomes and healthcare delivery.

7.2 Further Suggested Works

Future research on “Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach” can build upon this study by addressing several key areas. Firstly, expanding the dataset to include a larger and more diverse patient population would enhance the model's generalizability and robustness. Incorporating additional features, such as genetic information and lifestyle factors, could improve prediction accuracy. Additionally, developing real-time prediction systems integrated into electronic health records would facilitate timely and effective clinical interventions. Longitudinal studies to monitor the model's performance over time and its impact on patient outcomes would provide valuable feedback for continuous improvement. Collaboration with clinicians to refine and validate the models in practical settings is crucial for ensuring their applicability and effectiveness. Finally, addressing ethical considerations, including bias mitigation and ensuring patient data privacy, will be essential as these models are further developed and deployed in healthcare environments.

7.3 Limitations/ Conflict of Interests

This research on “Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach”, while promising, has several limitations. The primary dataset, collected via Google Forms, may not be fully representative of the broader patient population, potentially limiting the model's generalizability. The dataset's imbalance, with a high proportion of normal patients could affect the model's performance and lead to biased predictions. Additionally, the reliance on self-reported data introduces the possibility of inaccuracies and inconsistencies. While the selected algorithms (KNN, SVM, Random Forest, XGBoosting and Naive Bayes) are robust, each has inherent limitations, such as KNN's computational inefficiency with large datasets and XGBoosting's and Naive Bayesrisk of overfitting. The study also did not incorporate external validation, which is essential for assessing the model's applicability in real-world clinical settings. Furthermore, there may be potential conflicts of interest if researchers have affiliations or financial ties to organizations that could benefit from specific outcomes. Addressing these limitations in future research will be crucial for developing more accurate, reliable, and clinically useful predictive models for hemorrhoids.

REFERENCES

[1] Varut Lohsiriwat. Treatment of hemorrhoids: A coloproctologist's view. World J Gastroenterol 2015 August 21; 21(31): 9245-9252.

[2] G. G. Ravindranath, B. G. Rahul*. Prevalence and risk factors of hemorrhoids: a study in a semi-urban centre. Ravindranath GG et al. Int Surg J. 2018 Feb;5(2):496-499.

[3] Stefan Riss ,Friedrich Anton Weiser , Katrin Schwameis , Thomas Riss , Martina Mittlböck, Gottfried Steiner , Anton Stift. The prevalence of hemorrhoids in adults. Int J Colorectal Dis (2012) 27:215–220.

[4] Varut Lohsiriwat. Approach to Hemorrhoids. Curr Gastroenterol Rep (2013) 15:332.

[5] Tomas Poskus* , Zivile Sabonyte-Balsaitiene , Lina Jakubauskiene , Matas Jakubauskas , Ieva Stundiene , Gabija Barkauskaite , Mante Smigelskaite , Eugenijus Jasiunas , Diana Ramasauskaite , Kestutis Strupas and Grazina Drasutiene. Preventing hemorrhoids during pregnancy: a multicenter, randomized clinical trial. Poskus et al. BMC Pregnancy and Childbirth (2022) 22:374.

[6] Alberto Arezzo**, Vlasta Podzemny*, Mario Pescatori*. Surgical management of hemorrhoids: State of the art. Ann. Ital. Chir., 2011 82: 163-172.

[7] Felix Aigner, M.D., Gerd Bodner, M.D., Hannes Gruber, M.D., Friedrich Conrad, M.D., Helga Fritsch, M.D., Raimund Margreiter, M.D., Hugo Bonatti, M.D., Ph.D. The Vascular Nature of Hemorrhoids. J GASTROINTEST SURG 2006;10:1044–1050.

[8] Sheikh Firoj Kabir*, Debasish Das, Kazi Zana Alam, Mahbub Murshed, Din Mohammad. Frequency of Hemorrhoidal Complaints in a Real-Life Population and Possible Concomitance between Hemorrhoidal Disease and Chronic Venous Disease: Going Further in Our Understanding of Hemorrhoidal Disease. Surgical Science, 2021, 12, 319-331.

[9] Yago Ferreira Gomes1 , Filipe Alves Neto Verri2 , Dimas Betioli Ribeiro1# . Use of machine learning techniques for predicting the bearing capacity of piles. https://doi.org/10.28927/SR.2021.074921.

[10] M.D. Khan S, Pawlak SE, Eggenberger JC, Lee CS, Szilagy EJ, Wu JS, Margolin DA. Surgical treatment of hemorrhoids: prospective, randomized trial comparing closed excisional hemorrhoidectomy and the Harmonic Scalpel@ technique of excisional hemorrhoidectomy. Dis Colon Rectum 2001; 44:845-849.

[11] Wenzhuang Ma1 , Jintao Guo1 , Fan Yang1 , Christoph F. Dietrich2 , Siyu Sun1. Progress in endoscopic treatment of hemorrhoids. DOI: 10.2478/jtim-2020-0036.

[12] Kamal Ahmed Saeed, Faruk Hassan Faraj, Hardi Mohammed Dhahir*. Prolapsed hemorrhoids treatment options. https://doi.org/10.22317/imj.v4i2.815.

[13] Susan L. Gearhart, MD. Symptomatic Hemorrhoids. doi:10.1016/j.disamonth.2004.11.004 DM, November 2004.

[14] ROBERT D. MADOFF , JAMES W. FLESHMAN. American Gastroenterological Association Technical Review on the Diagnosis and Treatment of Hemorrhoids. GASTROENTEROLOGY 2004; 126:1463–1473. [15] P. J. Lunniss* and C. V. Mann. Classification of internal hemorrhoids: a discussion paper.

[15] W. H. F. THOMSON. The nature of hemorrhoids. Br. J. Surg. Vol. 62 (1975) 542-552.

[16] Steven D Wexner, MD, FACS, FRCS, FRCS (Ed), Fort Lauderdale, FL. The Quest for Painless Surgical Treatment of Hemorrhoids Continues. ISSN 1072-7515/01/$21.00 , PII S1072-7515(01)00997-8.

[17] Fabiola Delc6, M.D., Amnon Sonnenberg, M.D., M.Sc. Associations Between Hemorrhoids and Other Diagnoses.Dis Colon Rectum 1998;41:1534-1542.

[18] D.Bouchard*, L. Abramowitz†, A. Castinel‡, J. M. Suduca§, G. Staumont§, D. Soudan, F. Devulder**, F. Pigot*, M. Varastet†† and R. Ganansia¶. One-year outcome of haemorrhoidectomy: a prospective multicentre French study. doi:10.1111/codi.12090.

[19] Maia A. Qarabaki & Gela A. Mukhashavria & Gia G. Mukhashavria & Nodari G. Giorgadze. Circular vs. Three-Quadrant Hemorrhoidectomy for End-Stage Hemorrhoids: Short- and Long-Term Outcomes of a Prospective Randomized Trial. J Gastrointest Surg (2014) 18:808–815 DOI 10.1007/s11605-013-2424-x.

[20] Wei D, Jiang P, Gao R, Zhao Y. Prevention and treatment of anastomotic strictures after procedure for prolapse and hemorrhoids. Risk Management and Healthcare Policy. 2023 Dec 31:1351-7.
[21] Nassa YG, Danjuma A, Ayuba SB, Yahaya SA, Inusa B, Yakubu I. Prevalence and predictors of hemorrhoids among commercial motorcyclists in Kaduna state, Nigeria. World Journal of Preventive Medicine. 2016;4(1):1-4

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title:
PREDICTIVE-PROBABILITY ANALYSIS AND PREVENTIVE STRATEGIES FOR HEMORRHOIDS: A MACHINE LEARNING APPROACH

Student ID: [202-15-3791, 202-15-3804]

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a **Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids: A Machine Learning Approach** for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**

CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	For our research, we need to gather data from people & social media and understand how to create a machine learning-based model that addresses **K3** and **K4**	**Page no:** [2-3] **Section:** [1.4]
				It is necessary for our paper to include many components and we have designed a Google Form that includes **K5** and **K6**	**Page no:** [21-22] **Section:** [3.4]
				We have studied existing models with similar goals (**K8**).	**Page no:** [6-11] **Section:** [2.1, 2.2]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	We had trouble finding particular laws regulating the gathering of primary data from different individuals for our work.	**Page no:** [13-15] **Section:** [2.4,2.5]

3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	With so much primary data, it is difficult to come up with a definitive answer for our paper. We carefully considered a number of options and carried out a thorough investigation before selecting a particular algorithm.	**Page no:** [17-20] **Section:** [3.2,3.3]
4.	**EP4: Familiarity of Issues**	Yes	CO8	For our paper, we need to investigate a number of topics that we were not too familiar with in order to comprehend and analyze physical health and hemorrhoids. In particular, we need to determine the best way to figure hemorrhoid levels.	**Page no:** [21-22] **Section:** [3.4]
5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A
6.	**EP6: Extends of stakeholders involved and conflicting**	Yes	CO8	Our project included visiting and meeting with a doctor. The goal is to measure haemorrhoids based on a person's daily activities. This conference will help us determine the appropriate data for our analysis.	**Page no:** [2-3] **Section:** [1.4]
7.	**EP7: Interdependence**	Yes	CO5	Our paper consists of interdependent subsystems. These duties involve data collection, processing, machine learning model training, and prediction analysis.	**Page no:** [21-22] **Section:** [3.4]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	Our paper utilizes diverse resources such as high-performance computing infrastructure, machine learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements in hemorrhoid detection.	**Page no:** [24-27] **Section:** [4.1,4.2,4.3]
2.	**EA2: Level of interaction**	No		N/A	N/A
3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		This paper contributes to society by improving healthcare through advanced Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for patient data privacy.	**Page no:** [40-42] **Section:** [6.1,6.2,6.3,6.4]

5.	**EA-5: Familiarity**	Yes		This paper expands upon existing research by examining a novel approach in Predictive-Probability Analysis and Preventive Strategies for Hemorrhoids through machine learning, demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field.	**Page no:** [16-20] **Section:** [3.1,3.2]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This work tackles **CO4** through the integration of efficient project management and financial supervision, guaranteeing careful planning, resource distribution, and budget estimation for the best possible resource use during the course of the research process.	**Page no:** [21-22] **Section:** [3.4]
2	CO9	Yes	By putting patient privacy first, getting informed consent, and openly reporting the research process, the paper demonstrates adherence to ethical principles. It also ensures responsible knowledge dissemination and social well-being through the ethical application of cutting-edge healthcare technologies that abide by **CO9**.	**Page no:** [40-42] **Section:** [6.2,6.3]
3	CO10	No	N/A	N/A
4	CO12	Yes	The extensive data collection, rigorous statistical analysis, careful methodology development, and thorough experimental results and analysis all demonstrate the paper's dedication to continuous learning **(CO12)** and adaptation within the ever-evolving technological landscape. This shows a commitment to staying current and improving techniques to address contemporary challenges.	**Page no:** [28-39] **Section:** [5.2,5.3]

Plagiarism Report Checking for Final Defense v1 for Sajib and Tahsin

ORIGINALITY REPORT

18%	15%	9%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.science.gov Internet Source	2%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
3	Submitted to Daffodil International University Student Paper	1%
4	www.mdpi.com Internet Source	1%
5	fastercapital.com Internet Source	1%
6	"Optimized Predictive Models in Healthcare Using Machine Learning", Wiley, 2024 Publication	1%
7	www.frontiersin.org Internet Source	1%
8	Submitted to Liverpool John Moores University Student Paper	1%